DEVELOPMENT OF A NEW DISTRIBUTED WATER QUANTITY AND QUALITY MODEL COUPLED WITH REMOTE SENSING AND GEOGRAPHIC INFORMATION SYSTEMS (GIS) AND ITS APPLICATION IN A SMALL WATERSHED

(リモートセンシングおよび地理情報システム（GIS）と連携した新しい分布型水質水量モデルの開発とその小流域への適用)

BINAYA RAJ SHIVAKOTI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Engineering
Department of Urban and Environmental Engineering
Graduate School of Engineering

Kyoto University
Kyoto, Japan
September 2007

ABSTRACT

Development of a New Distributed Water Quantity and Quality Model Coupled with Remote Sensing and Geographic Information Systems (GIS) and its Application in a Small Watershed

BINAYA RAJ SHIVAKOTI

Understanding river water quantity and quality variation is one of the fundamental requirements for the integrated watershed management. Monitoring is usually preferred to examine and understand the river water quantity and quality, especially focusing on pre-specified objectives. Although monitoring is invaluable in many instances, it is of less use to forecast the foreseeable changes, especially, for the long-term prediction that is usually required by the decision-makers. Therefore, for the decision-making, modeling is widely practiced. Due to the limited understanding of hydrological processes inside a watershed, models often fail to estimate properly, which in worst case could often mislead the targeted plans. Among several aspects, spatial variability such as land cover, topography, soil, geology is believed to affect the overall performance of the model. Such thought lead to the concept of distributed models that were supposed to represent spatial variability through modeling specific variations inside the watershed by using several representative units or grids. In that meaning, distributed models required to identify and assign the values of its parameters to represent the physical processes defined by the governing equations for each grid. Due to the unavailability of required spatial information at appropriate grid sizes, even physically based and conceptually sound distributed models fail to estimate properly thereby offsetting the credibility of distributed models. Therefore, in this study, we set a major objective to develop a new distributed water quantity and water quality model to address some of the stated issues. Major emphasis was given to conceptually sound but simple structure of the model. In addition to that, model aimed to utilize the potential of recent advances in spatial information, such as remote sensing and GIS, to generate and process the spatial data, and to determine the values of its essential parameters. The approach was expected to provide an example that the complexity of the model should be preferred only if the defined processes could be ascertained within some reasonable limit.

At the initial stage, several spatial data were collected from different sources and they were processed into raster format, which was one of the essential requirements for the distributed model. Analysis of spatial database indicated that the watershed was characterized by forested parts in the hills, and densely populated urban areas in plains. Rainfall occurred quite frequently but they were of short duration. Besides constructing spatial database, several water quantity and quality surveys were also conducted at different spatial and temporal conditions from 2000 to 2006. The data were mainly used to understand variation patterns of water quantity and quality at both spatial and temporal conditions. Later on, some of the data were also used for the verification of model in study area. 28 water quality indices (WQIs) were observed for each observation, which were mainly utilized to understand the overall variation pattern of river water quality. Initial analysis of flow rate condition of the river showed that the rainfall-runoff responses were quite rapid after the rainfall but such effect appear for very short duration (< 2 days). Then, analysis of variance (ANOVA) and two multivariate analysis techniques (MVA), namely, principle component analysis (PCA) and cluster analysis (CA) were used to explore effectively the river water quality datasets. Analysis showed that the observed covariation among majority of WQIs could be due to the inter-linkages among rainfall pattern, atmospheric deposition of acidic ions, soil and geology of dominant forest areas, topography, and climatic conditions. The identified pattern indicated that there could be close relationship between the biogeochemical processes in the forest areas with both river water quantity and quality variation.

A new distributed water quantity and quality model was developed especially focusing on the biophysical characteristics of the watershed. Basic structure of the model was similar to the concept of lumped tank model, which was often credited for its simple and sound conceptual structure. Two storey tanks were conceptualized for each grid, but model also took into consideration of drainage channels in urban areas and natural river channels as rapidly conveying structures. Besides, the model considered all major aspects affecting the estimation of water quantity, such as interception of the rainfall, evapotranspiration loss, surface runoff, sub-surface runoff, and ground water runoff. Compared with the original tank model, major emphasis was given to assign the values major parameters, such as coefficients and storage heights of the outlets, by relating them with the hilly topography of the study area and the variation in land cover, soil, and

geology. The model was further integrated with water quality component, which was based on two fundamental assumptions of build-up and wash-off of the WQIs in the environment. Build-up was based on the land cover type and population, while wash off was based on the estimated runoff volume.

Remote sensing and GIS techniques were used to assist in the modeling process. At first, remote sensing was mainly focused in the classification of land cover by utilizing seasonal Landsat ETM+ images. In addition to urban and vegetated urban categories, four major forest categories (shaded, deciduous, mixed, and evergreen) were identified. Then leaf area index (*Lai*) was determined for each vegetation category. *Lai* was mainly used to determine the rainfall interception by the canopy in the forest areas. In this study, forest areas showed the capacity to intercept as high as 1.2 mm of rainfall, which could be quite important during smaller rainfall events. Remote sensing was further used to determine the transpiration coefficient of the vegetations, which was a major requirement for the estimation of evapotranspiration (*Et*) loss by the FAO Penman- Monteith method used in the model simulation. *Et* was estimated even reached more than 4 mm/d in summer months, but it was relatively lower (< 2 mm/d) in the winter months. These facts suggested that consideration of both interception and *Et* loss in a forested watershed could have significant influence on the estimation of flow rates by the model.

At the final stage, model was applied in the study area. Mainly three approaches were considered to assess the estimation by the model. First was conventional approach in which comparison between the observed and estimated data were done considering different spatial and temporal contexts. Assigned values of the parameters gave satisfactory prediction for both water quantity and quality for the selected grid size of 50 m in which the relative error was usually less than 1. The second approach evaluated the model by considering different scale of the grids ranging from 100m to 500m. It was observed that grid resizing usually affected the basin attributed such as slope, outlet height, drainage characteristics following nearly proportionate pattern than other categorical variables such as land cover or geology. Usually same parameter values gave very different prediction level for both magnitude and shape of the hydrographs (or pollutographs), in which increasing grid size was accompanied by the increasing peak event estimation or overall error. The effects were further assessed by changing the values of key parameters for each grid size targeting the minimum differences between the observed and

estimated values. Interestingly, the parameters also showed some identifiable (increasing or decreasing) trend with the change in grid size. Particularly, due to the direct effect of predicted runoff on the reference WQIs, its showed more complex variation pattern at different grid sizes. Overall assessment of the distributed model indicated that the model was quite sensitive to the selection of key parameters for different grid sizes. It indicated that the values of calibrated parameters might not give stable result if the scale of input data were changed. It could further indicate that the choice of grid size should be assessed before the actual application of the model considering the spatial variability of the watershed. In the third approach, model was utilized to estimate at different scenarios, namely, rainfall variation and land cover changes. The differences in the estimated results could indicate that the model could be available for the watershed management at different runoff and land cover scenarios in future.

ACKNOWLEDGEMENTS

It is a moment of pleasure to extend profound gratitude and sincere thanks to my adviser Prof. Shigeo FUJII for giving an opportunity to study at Kyoto University, and providing a wonderful chance to observe and experience interesting lifestyle of Japanese people. I appreciate him for sharing his experiences, for helpful suggestions, critical comments, and showing invaluable moral support during the whole study period that made me capable to prepare this dissertation.

I am equally obliged towards Prof. Hiroaki TANAKA and Prof. Yoshihisa SHIMIZU for their constant encouragements, invaluable advises, and kind help during my study at Research Center for Environmental Quality Management (RCEQM), Kyoto University.

Special appreciation also goes towards Ministry of Education, Culture, Sports, Science and Technology, Government of Japan, for all financial supports to study at Kyoto University.

Special thanks also goes to Dr. Shuhei TANAKA for sharing his invaluable experiences, showing his great concerns about my study, and extending his kindness time and often. I am also pleased to acknowledge Dr. Naoyuki YAMASHITA for his moral support.

I would like express my thankfulness to Ms. Kinuyo FUKUNAGA and Ms. Kazumi HATTORI for their kindness and concerns. I also consider my pleasure to acknowledge Dr. Kazuhiro IKEDA and Dr. Taketoshi KUSAKABE for the extended help during the whole study period. Besides, I appreciate all RCEQM members for the friendly and cordial environment.

The study will not be complete without acknowledging former students of Prof. FUJII laboratory, especially, Hirotaka IHARA, Masashi MORYA, and Kazuaki SHICHI for setting up the foundation of this study. I specially thank Kazuaki SHICHI for helping me to catch-up with the research and teaching me to use several analytical equipments during initial days. I thank Ms. Naoko MORITANI for making my life easier and supporting my research work. I understand, it is not necessary to mention Ms. Nguyen Pham Hong Lien, who constantly inspire and support me as a good friend. I consider Dr. Suwanna KITPATI and Dr. Piyaporn SONGPRASERT as well wishers. I thank all other members of Prof. FUJII lab for their moral support.

Besides, I would also like to acknowledge Prof. Ganesh Shivakoti of Asian Institute of Technology and my former advisor Dr. Michael Albert Zoebisch. Finally, it is time to mention about my parents who are always looking forward to my each step.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION AND STRUCTURE OF THE DISSERTATION

1.1 Background

There is increasing awareness that water will be the most critical natural resource in future. Water scarcity is increasing worldwide and it has put pressure on the existing water recourses due to the growing demands in several sectors such as for drinking purpose, industrial use, agricultural production, hydropower generation, and so on. Besides, point and non-point sources pollution have contributed to the deterioration of water quality. Efficient water resources management would be required in order to meet the growing demand of water as well as to deal with several pollution problems in future.

Among water resources, river systems are of primary importance due to their highly mobile state and far reaching consequences. Rivers are the habitats of many flora and fauna, important source of freshwater to downstream inhabitants and industries, principle carrier of nutrients or pollutants, and sometimes cause destruction to life and economy due to the flooding.

State of a river system is directly or indirectly affected by various natural (physical, chemical, biological) as well as anthropogenic factors inside the watershed. Most of these factors are common to all watersheds, but each watershed is characterized by the variability in the composition of all governing factors, such as hydrology, terrain, geology, soil, vegetations, land uses and so on. Such varying composition makes one watershed unique than the others and shows different degree of rainfall-runoff variability. Monitoring programs are usually preferred to understand the rainfall runoff variability and resulting water quality of a river system. In monitoring programs, relevant data are collected and later on analyzed to sort out the possible factors governing the water quality variations. Such monitoring generally gives an idea about the state of the water quality possibly valid for a limited range of time or within the pre-specified objectives.

To overcome the limitations of monitoring programs, some sorts of decision support tools are required, especially, for the assessment of future trends of both water quantity and quality. Modeling approaches are widely preferred to understand the likely changes that could occur in future. In most of the cases, models are the only cost effective tools available for the decision makers. In modeling, water quantity and quality variations are estimated by defining different governing processes inside the watershed. Based on the estimated output appropriate management decisions are proposed.

Effectiveness of the models depends on the estimation by it on different rainfall runoff conditions. In the initial days, lumped or black box types models were usually preferred. It was soon realized that water quantity and quality estimated by the model were affected directly or indirectly by the major watershed variables such as land use, vegetations, soil, geology, terrain *etc*. Distributed models were then developed to incorporate the spatial variability in the modeling process. One of the major requirements for distributed modeling was to design a system that was capable to deal with the spatial information. Next requirement was to construct distributed databases that could be used directly or indirectly during the parameterization of the model. However, majority of available distributed models are quite complex in terms of modeling process that require to deal with huge amount of spatial information. Many of the parameters' values required by the distributed models could not be determined with certainty. Especially, introduced uncertainty by applying unknown values of parameters at different locations could restrict the purpose and potential of distributed models, which is also one of the biggest issues in distributed modeling.

Recently, advances in computing capability, geographic information system (GIS), and remote sensing techniques have gained wider attention due to their potential to improve our understanding of spatial variability in a watershed and provide ways for the parameterization of distributed models. Specifically:

1) Computing capability has made possible to store and analyze large volume of databases which is primarily required by any distributed models;
2) Geographic information system (GIS) has its capability to store, display, and analyze the spatial datasets;
3) Advances in remote sensing techniques, which has the potential to generate the data of larger spatial interest;

Utilization of these techniques could possibly improve our understanding of the factors governing the variation of both water quantity and quality and could also provide a potential option to enhance the effectiveness of distributed modeling in future.

1.2 Objectives

The main objective of this study was to develop a new distributed water quantity and quality model that would be simple in its design but could utilize the remotely sensed data and GIS to assign the distributed values of important parameters. The model was mainly focused for the application in the small, hilly, and forested watersheds. To fulfill the major objective, several specific objectives were stated as:

1) To construct the various types of spatial databases of the study area and understand its general characteristics;
2) To collect river water quantity and quality data through several surveys and to explore their spatial and temporal variation patterns;
3) Develop a new distributed water quantity and quality model based on the concept of lumped tank model;
4) Process the remotely sensed data and elevation data to derive the distributed values of important parameters required by the model;
5) Assess the model performance for its applicability in the study area;

1.3 Structure

This dissertation consists of seven main chapters and one more chapter for conclusions and recommendations. Conceptual overview of the whole study is shown by the flow chart in **Fig. 1.1**.

The theme of each chapter is summarized as follows:

Chapter I Background, objectives, and structure of this dissertation.

Chapter II Review of relevant literatures.

Chapter III Description of study area and spatial database construction.

Chapter IV Data collection of model verification and identification of river water quantity and quality variations in the study area.

Chapter V Concept and design of a new distributed water quantity and quality model based on the conceptual framework of tank model.

Chapter VI Describes the processing methods of satellite images and digital elevation model (DEM) to derive the values of important parameters of the model.

Chapter VII Assess the applicability of the model in the study area through verification of estimated results with respect to the observed data, through analysis of the effects of different grid sizes on model estimation, and through the evaluation at different scenarios.

Chapter VIII Conclude the main findings and recommend for further research.

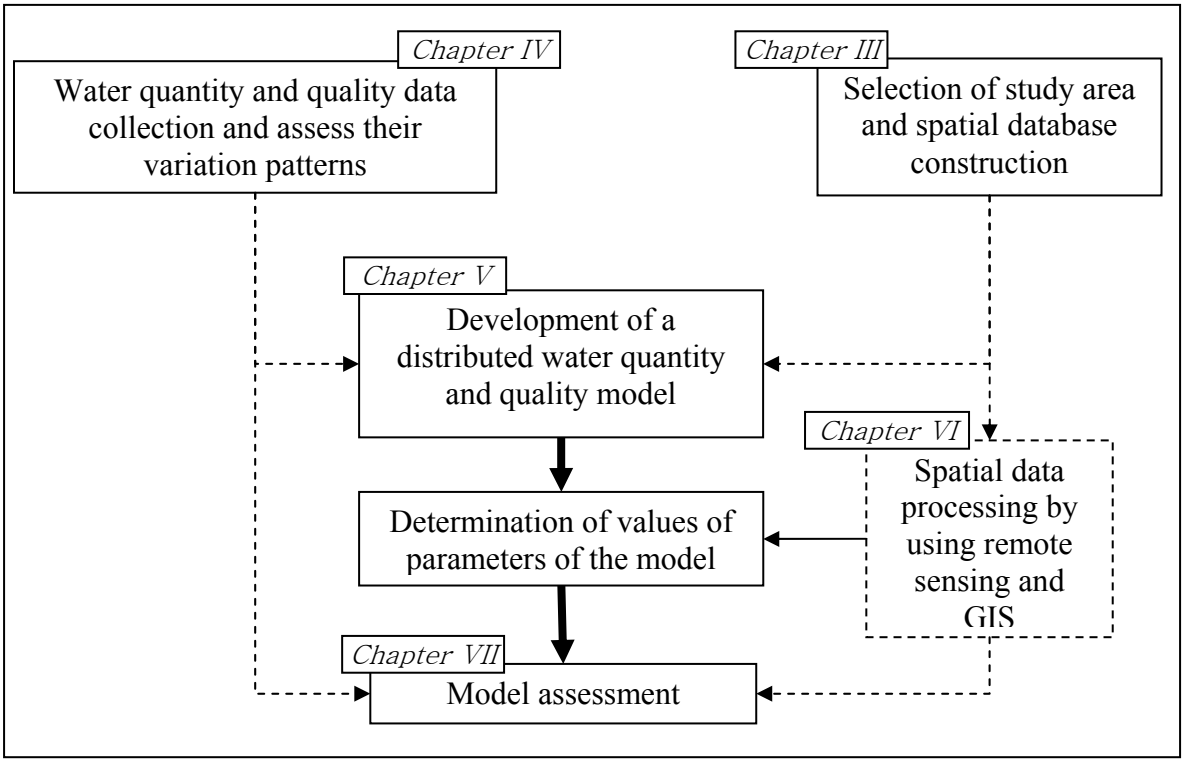


Figure 1.1 Flow diagram showing the structure and scope of this study

CHAPTER II

LITERATURE REVIEW

2.1 Introduction

This chapter aims at elucidating the advances and issues related with this study with the help of published literatures. In the initial part, common statistical water quality (and quantity) analysis methods relevant to this study are described. Characteristics of watersheds in Japan and their contribution to the variation of river water quantity and quality will be assessed, especially focusing on the role of forest areas. Then brief introduction of river water quantity and quality modeling will be given, especially focusing on significance of distributed modeling. Major applications of remote sensing and GIS in distributed modeling will be shown. Finally, the general methods and issues related with the assessment of the prediction by distributed model will be given in which specific focus would be given to the issues of scale.

2.2 Water quality (or quantity) data analysis

Water quality reflects the composition of water as affected by nature and human cultural activities expressed in terms of both measurable quantities and narrative statements (Novotny, 2003). Quantity and quality of water body is due to the interaction of three major components of its watershed: hydrology, physico-chemistry, and biology. The degree of interaction of these components is determined by both natural and anthropogenic factors. Monitoring programs are usually conducted to assess the effect of natural and anthropogenic factors on the water quality. However, data obtained from such monitoring programs are of limited significance unless they pass through some sort of analysis process. In that meaning, analysis of water quality data is very important, and usually constitutes the last step of any assessment process, because it shows how successful the monitoring has been in attaining objectives of assessment (Demayo and Steel, 1996).

Statistical techniques are usually preferred in the analysis of water quality data beca-

use they offer efficient ways to explore the observed data. Statistical analysis techniques could be broadly categorized into two main groups on the basis of the dimensions of variables: 1) univariate (or bivariate); and 2) multivariate. Univariate (or bivariate) statistical techniques are the simplest type and are useful for the preliminary assessment of any type of water quality data. In many cases, especially those dealing with a few number of specific water quality indices, univariate (or bivariate) constitute the major technique to analyze the water quality data. When the number of interest water quality variables is large and the volume of collected data over long monitoring periods are very huge, meaningful interpretation of the dataset becomes very difficult by using simple univariate (or bivariate) statistical measures (Dixon and Chiswell, 1996). Multivariate analysis (MVA) techniques are appropriate to analyze such multidimensional water quality dataset together. MVA techniques are quite useful for the assessment of large and complex databases in order to get better understanding of surface water quality, designing of sampling and analytical protocol, and effective management of surface water quality (Simeonov *et al.* 2003; Singh *et al*., 2005). MVA are widely used as pattern recognition techniques in the assessment and evaluation of surface water quality, especially, due to the spatial and temporal variation caused by the natural and anthropogenic factors (Singh *et al*., 2004; Wunderlin *et al.*, 2001; Vega *et al.*, 1998; Perona *et al.*, 1999; Bengraine and Marhaba, 2003). Among the MVA techniques, two pattern recognition techniques, namely, cluster analysis and principle component analysis are widely used in water quality data analysis because both are unbiased methods, *i.e.*, they do not require any presumption about the distribution of data. A brief overview of analysis of variance, cluster analysis, and principle component analysis is given below because these techniques constitute the major portion of water quality data analysis in this study.

2.2.1 Analysis of variance (ANOVA)

ANOVA is a powerful statistical methodology in which the total variation in a measured response is partitioned into components that can be attributed to recognizable sources of variation (Milton and Arnold, 2000). ANOVA can be one-way, two-ways or multivariate (MANOVA). One-way ANOVA is used when one factor is being tested for different levels or treatments, such as concentration differences of a water quality item observed at different places. Two-ways layout ANOVA, on the other hand, is more

suitable if the effect of two important factors on the total variance of water quality data is sought. Two-ways layout ANOVA is a powerful technique to sort out the factors for the variance, if the variation caused by the error is minimal. For example, Vega *et al.*, (1998) has used ANOVA of rotated principle components, in which the mineral content in river showed higher seasonal variation and the organic pollutant showed site-specific variations. ANOVA is well documented in all good statistics textbooks, so further details could be accessed from any (*e.g.*, Milton and Arnold, 2000).

2.2.2 Cluster analysis (CA)

CA is a set of tools for building groups (or clusters) from multivariate data objects that aim to construct groups with homogenous properties out of heterogeneous large samples (Hardle and Simar, 2003). In CA, target is to maximize intra-cluster similarity but to minimize the inter-cluster similarity. There are two fundamental steps in CA:

A) Choice of a measure of similarity:

Measure of similarity is usually determined by the distance between the variables. Common methods of distance measurements in CA are Euclidean, Manhattan, Mahanalobis *etc*. Among them Euclidean distance is the simplest and widely used distance measure in water quality data analysis. Euclidean is simply the geometrical distance between two variables. The squared Euclidean distance between any two points (*i,j*) in *p* dimension is calculated as:

$$b_{i,j} = \sum_{k=1}^{p} (x_{ik} - x_{jk})^2 \qquad \textbf{2.1}$$

$$B(n \times n) = \begin{pmatrix} b_{11} & b_{12} & \cdots & \cdots & b_{1n} \\ b_{21} & b_{22} & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ b_{n1} & \cdots & \cdots & \cdots & b_{nn} \end{pmatrix} \qquad \textbf{2.2}$$

Where x_{ik} and x_{jk} are the value of k^{th} variable for objects *i* and *j* respectively. Then for *n* objects or measurements, a $n \times n$ dimensional squared Euclidean distance matrix (D) could be constructed (Eq. 2.2). Based on the $B(n \times n)$ matrix, clustering criteria are developed.

B) Choice of group-building algorithm or clustering criteria

The common clustering criteria, according to Scott and Clarke (2000) are:

a) Nearest neighbor (single linkage): distance between two groups is the furthest distance between any point in the first group and any point in the second group;

b) Furthest neighbor (complete linkage): distance between two groups is the shortest distance between any point in the first group and any point in the second group;

c) Group average: distance between two groups is the average of all possible distances between any point in the first group and any point in the second group;

d) Median: distance between two groups is the median of all possible distances between any point in the first group and any point in the second group;

e) Centroid: distance between two groups is distance between the means of the two groups;

f) Minimum variance (Wards's method): joins the two groups for which the increase in overall within cluster variance is least;

CA can be broadly divided into non-hierarchical and hierarchical techniques. In non-hierarchical technique (*e.g.*, *K*-means clustering) number of clusters are chosen *a priori* and based on preset similarity or dissimilarity criteria clustering is done. In hierarchical CA, clusters are grouped as hierarchy or tree like structures. The hierarchy could be agglomerative (from small clusters to big one) or divisive (from big cluster to small clusters). Agglomerative is widely preferred to divisive because of less computational requirement (Scott and Clarke, 2000).

2.2.3 Principle component analysis (PCA)

PCA was the first ordination technique to be developed and still the most used method of ordination (Scott and Clarke, 2000). In PCA linear combination of original variables are sought such that the combination embody as much as possible variance of whole dataset. The linear combination could be written as:

$$z_{ik} = a_{i1}x_{1k} + a_{i2}x_{2k} + \cdots + a_{ij}x_{jk} + \cdots + a_{im}x_{mk} \qquad \textbf{2.3}$$

Where, z_{ik} is the score for i^{th} principle component (PC) and k^{th} sample, a_{ij} is the component coefficient for i^{th} *PC* and j^{th} variable, x_{jk} is the observed value of j^{th} variable and k^{th} sample,

and *m* is the total number of variable. PCs are arranged in terms of descending variance and they are uncorrelated among each other.

$$\left.\begin{array}{l} Var(PC_1) > Var(PC_2) > \cdots\cdots > Var(PC_i) > \cdots\cdots > Var(PC_m) \\ Correlation(PC_i, PC_n) = 0 \qquad \{i \neq n \end{array}\right\} \qquad \textbf{2.4}$$

Where *n* is any *PC* other than i^{th} *PC*. Because of these two main characteristics of *PC*, in ideal case, first few PCs are enough to explain almost all of the variance of the dataset. Therefore, two main uses of PCA are to reduce the dimension of multidimensional dataset and to sort out the covariance pattern among the interest variables.

2.3 River water quantity and quality variations in Japanese watersheds

2.3.1 Characteristics of Japanese watersheds

Japan is a mountainous, humid, and forested country, with its people concentrated in densely populated urban areas along the coast and on alluvial plains. Mountain constitutes nearly 72 % of the land, of which 65% have slopes steeper than 14% (Yoshimura *et al*., 2005). These hilly areas of the Japan are mostly characterized by forest cover and agriculture. Forest accounts nearly 66% of national land resources of Japan, agriculture constitutes 12.5 % that are mostly dominated by paddy cultivation, and build-up urban area constitutes only 8.3% (SBJ, 2007). Average population density of Japan is 343 capita/km^2, but in urban parts, it generally exceeds 5000 capita/km^2.

Geologically, Japan is young country with active mountain formations, frequent earthquakes, and numerous volcanoes. Geology in mountainous areas are composed of sedimentary rocks, sandstone, chert, and limestones, where lowland areas consists of alluvial fans, flood plains, and deltas that are underlain by coarse sediments and volcanic ashes. Due to relatively young geology, and frequent land slides occurring in the steep slopes the soil profile are reported to be poorly developed or at initial stage of development. Most of the soils in the hilly forest areas belong to brown forest soils (Cambisols (FAO) / Inceptisols (USDA)) that are characterized by low amount of weathered materials and higher base saturation (Ohte *et al.* 2001a; Takahashi *et al*., 2001).

Japan experience seasonal change in the climate showing both warm summer and cool winter. Northern parts are more temperate than the central or southern parts of Japan.

Annual precipitation at main regions of Japan in 2004 was in the range of 1063-2427 mm. Most of the watersheds in Japan are frequently affected by very heavy storms, but such storms are usually of shorter duration. Heavy precipitation is usually caused by rainy season in June and July (except at north), typhoons from the Pacific Ocean in September and October, and winter snowfall on the north side of islands.

2.3.2 River water quantity and quality

Rivers in Japan are short, steep, and often show flashy flow regimes generally lasting less than two days. Especially, flashy flow regimes are caused by the combined influence of steep topography, and frequent, short duration but intense rainfall events. Such type of river flow often aggravate fluvial processes causing huge loss of sediments (Oguchi *et al*., 2001) and also are responsible for the wash-off of the nutrients and pollutants from the upstream areas.

Availability of the nutrients and pollutants for the wash-off are determined by the complex biogeochemical processes inside the watersheds, which are generally specific to land-use characteristics. The major land use types, namely, forest, agriculture, and urban are of primary importance because river water quality is determined by the interaction of natural as well as human influences in these land uses. Agriculture in Japan is mainly rice-based cultivation, while vegetables, fruit orchards, and animal farming also constitute important part. Agriculture as a whole is considered as a major non-point source of nutrients such as nitrogen and phosphorus to the river and other water bodies. Impact of agriculture on the water bodies are often concerned in Japan (Sato *et al*., 2004; Kato, 2005). Urban areas are point sources that are often associated with range of domestic and industrial wastes. Especially after introduction of environmental pollution prevention act 1970, contribution of pollutants to the river from point sources was significantly checked (Yoshimura *et al*., 2005).

Forest is the largest land cover in most of the Japanese watersheds. In recent days, forests were considered important from viewpoint of soil and water conservation rather than timber production (Murakami *et al*., 2000). The main reason for that was the increasing demand of fresh water, drought, flood damage, and decreasing price of timber. In Japan, enormous areas shared by forests could exert dominant effect on both river water quantity and quality. According to Ohte *et al*. (2001a), one of the major themes of future

hydrobiogeochemisty research on Japanese forest should be the evaluation of stream water chemistry as an output from the forest ecosystem. Understanding of major processes and factors affecting the biogeochemistry of the forest ecosystem are quite important for the sustainability of river water quality and forest hydrology.

Acid rain is a serious environmental problem in Japan and its impacts on the forest ecosystem and river water quality are widely concerned (MOE, 2003). Associated issues related with the acid rain are the rapid weathering of the rocks, base-cation exchange in the forest soils, and nitrogen saturation in the forest. Acid rain in Japan is found to be neutralized by the soil and bedrock so that the receiving water body usually do not result acidity, where as in North American or European countries acidity is usually observed as a result of acid rain (Asano and Uchida, 2005; Ikeda and Miyanaga, 1995). According to Asano and Uchida (2005), role of soil in acid neutralization is when the soil is not leached of base cations. When the overlying soil is leached of base-cations, the role of bedrock in acid neutralization is very significant. Chemical weathering of primary minerals could be the main factor to support the neutralization of acid rain (Ikeda and Miyanaga, 1995). Although effect of acid rain is not significant to create acidity in the river or fresh water body, it is however evident that neutralization process will continually supply the primary minerals. Anazawa and Ohmori (2001) with the help of multivariate analysis has concluded that major ions of acid rain rapidly leached out major cations (Na, K, Ca, Mg) from the summit area of Norikura. Similarly, Ohuri and Mitchell (1998) have mentioned that major ions in the streams in forested areas are dominated by the IC, HCO_3^- , Ca, and NO_3. Nakamura *et al*. (1984) had mentioned that carbonic acid and sulfuric acid were the main cause of high weathering of rocky minerals in Tenryu river basin, where higher concentration of HCO_3^-, SO_4, Na, Mg, Ca, and Mg was observed.

Frequency and intensity of precipitation, warm and humid climate in summer, steep topography, forest ecosystem, acid rain, nitrogen saturation, roles of soil and geology in acid neutralization are interlinked to the resulting river water quantity and quality. Their combined effect on the river water quality is very important for the assessment of risk of acidification in water bodies or eutrophication (Ohte *et al.*, 2001a).

2.4 Modeling river water quantity and quality

2.4.1 Modeling concepts and types of models

Modeling of river water quantity or quality is an attempt to predict its variation at different spatial and temporal conditions. Models are means of extrapolating the available measurements, in space and time, to access the likely impact of future changes inside a watershed that will hopefully be helpful in decision-making (Beven, 2001a). Models are probably the widely used decision-making tools since it is not possible with present state of knowledge to measure and understand all the processes occurring inside the watershed that could affect hydrology and water quality. Models could be useful in several instances such as runoff predictions (flooding, effect of land use change, effect of climate change), environmental issues (water scarcity, soil erosion, non-point pollution, acidification, deterioration of aquatic systems, eutrophication), groundwater conditions, ecological simulations and so on.

In practical terms, the initial aim of water quality modeling is to determine the fate of water inside the watershed at the given condition (Martin and McCutcheon, 1999). Later on, it also provides the basis to determine the fate of nutrients or pollutant that travel along with the moving water. In fact, majority of river water quality models are the integrated from the hydrological models as reported by Arheimer and Olsson (2003). All models are designed on the basis of some conceptual framework that describes the hydrological processes occurring in any watershed. Although there are different ways to classify the models (*e.g.*, Jorgensen, 1989, Singh, 1995; Refsgaard, 1996), models could be classified on two broad bases:

A) Stochastic or deterministic

Stochastic models are characterized by the predictions that show some variations or randomness, which are assumed the function of some type of probability distribution. Such models contain stochastic input disturbances and random measurement errors (Jorgensen, 1989). Deterministic models, on the other hand, are guided by perfect knowledge of the system and have only one output without statistical distribution. Both stochastic input disturbances and random measurement error are zero in deterministic models.

Majority of models used in hydrology are deterministic in nature, but there are no clear-cut distinction between deterministic and stochastic models (Beven, 2001a). Some models are deterministic in nature but could add stochastic error model in it or some models use a probability distribution function of state variables but make prediction in

deterministic way. According to Beven (2001a), irrespective of complexity in the modeling processes, the rule-of-thumb is that model having single output is deterministic and the model showing dispersion on the estimated output is stochastic. Especially, when the randomness in the estimated results were to occur by the variation in the parameter values, it is usual practice to assess the model performance by stochastic way such as Monte-Carlo type simulation or application of generalized likelihood uncertainty estimation (GLUE) methodology.

B) Lumped or distributed

Lumped or distributed models are characterized by the representation of spatial differences in the modeling process. If whole watershed is a single unit and at the given condition models variables are static over space and time, the model type is lumped. All the processes described in a lumped model therefore represent the average response from the whole unit.

Distributed models are characterized by variables and processes distributed over the space (and time) represented in terms of units or grids. Although distributed models are often termed as physically-based, at the scale of individual grid the model is essentially a lumped model. Therefore, for a watershed, distributed model is equivalent to lumped type when grid scale becomes coarser, but it approaches to more physically based structure when grid scale becomes finer.

2.4.2 River water quality modeling

Water is not only the principle agent for the transport of nutrients or pollutants but also the major medium for the transformation of nutrients or pollutants through various biochemical reactions (*i.e.*, kinetics). Therefore, primary tasks of all river water quality modeling are to estimate the water quantity at given space and time and then to integrate the transport and kinetics behavior of nutrients or pollutants with the hydrological processes (Martin and McCutcheon, 1999; Arheimer and Olsson, 2003). In most basic form, water quality is generally modeled by considering two main processes. First process is the natural or artificial buildup of the pollutant over a given time period which is determined by the rate of production and decay in a given land cover type. Next process is the wash-off of the pollutants by the surface runoff or through leachate to the underground.

Movement of nutrient or pollutants in a river system could be explained by two phenomenons, namely, advection and diffusion (Chapra, 1997). Advection is the organized movement of molecules while diffusion is due to the random movement of the molecules of water. Diffusion could be at microscopic scale, i.e., molecular diffusion, caused by the Brownian motion of water molecules, or it occurs due to the mixing by turbulent eddies or whirls inside the water bodies, *i.e.*, turbulent diffusion. When the mixing is due to the velocity of flowing water it is termed as dispersion, which has higher diffusion coefficient than turbulent diffusion. In natural rivers and streams, mixing is caused by both turbulent diffusion and dispersion (Chapra, 1997).

Both movement and kinetics are specific to each nutrient or pollutant type, and need to be dealt separately in the model. Specific behavior of each nutrient or pollutant often complicates the modeling of water quality as compared to the hydrological modeling. Movement is determined by the individual pollutants resistance to the flowing water or rainfall. Particularly, the particulate associated pollutants show different degree of mobility as compared with the pollutants that are easily dissolved with the water. Kinetics is important especially when the production and decay of particular pollutants are significant for a given time period. One of such example is the modeling of nitrogen, which is often characterized by the transformation into different oxidation states by nitrification or denitrification processes. QUAL2K (Chapra and Pelletier, 2003) and SWAT (Arnold *et al*., 1998) are the examples of popular water quality models which are designed to model specific behavior of water quality indices such as DO, ammonium, phosphate, algal growth, BOD, toxic substances, pesticides and so on, in the natural environment.

2.5 Distributed modeling

As mentioned earlier, distributed models aim to integrate the spatial variability of a watershed. After the introduction of the first blueprint of a distributed hydrological model (Freeze and Harlan, 1969), major attraction towards distributed modeling was mainly due to the enhanced computing capability, GIS techniques, and easier availability of distributed data. The fact could be realized from increasing number of distributed models, such as MIKE-SHE (DHI, 2004), IHDM (Rogers et al., 1985), TOPMODEL (Beven, 2001), AGNPS (Young *et al*. 1989), WATFLOOD (Kouwen *et al*., 1993; Leon *et al*., 2001), SMR (Frankenberger *et al*.,1999). Apart from possibility of applying distributed modeling in

recent days, it is in fact the only practical option that will be needed to address the growing concerns on various environmental and water resource problems such as non-point pollution problem, to assess risk of soil erosion, to predict the impact of land use changes on water quantity and quality and so on (Beven, 2001b; Smith *et al*., 2004). WATFLOOD (Leon *et al*., 2001), PROW (Zhang and Yamada, 1996), AGNPS, ANSWERS-2000, SWIM models are some examples of distributed models that are especially designed to deal with non-point pollution at different scales.

All of the stated distributed models are in fact developed to improve the estimation for both general to specific scenarios. Due to the limited understanding of the major hydrological and biophysical processes in a watershed scale, all these models are essentially based on very similar fundamental principles. Such conditions often caused the ineffectiveness of even physically based distributed models. One of the central ideas of distributed modeling was that all the processes at finer scale could be physically translated into the output such that uncertainties through the calibration were almost negligible. For instance, in theory MIKE SHE, IHDM, or TOPMODEL has physically based structure and they do not need preliminary calibration because model parameters have clear physical meaning, but in practice calibration are often required to determine the parameter values. It could be due to the nature of governing equations that have physical interpretation, but some of its parameters values could not be physically quantified such as hydraulic conductivity or roughness coefficients of given land use type. According to Beven (2001a), nonlinearity, scale, equifinality, uniqueness, and uncertainty are the major issues that, in the reality, often contradicts between the physical meaning of distributed models and actual prediction. Nonlinearity is inherent to the distributed modeling because responses due to spatial variability often follow nonlinear pattern, which could not be expressed by the governing equations at the present state of knowledge. Scale determines how finer details of watersheds are represented by the distributed model. In that sense, differences caused by the scale of model is sure to restrict the model estimation due to the modification in the information. Equifinality states that, due to the lack of understanding of (non-linear) hydrological processes at the watershed scale, different models or set of parameters could give equally valid prediction in a basin such that importance of single parameter or single model structure could not be easily assessed. In fact, some degree of calibration is always required due to the equfinality problem. Even with very perfect model (*i.e.*, physically

based), uniqueness of the place restricts the estimation of true parameter values at a given scale. Therefore, availability of limited measurement of distributed parameters often restricts the identification of optimal models leading to equifinality problem. Finally, stated problems of nonlinearity, scale, uniqueness, and equifinality always introduce estimation uncertainty. Although stated problems are interconnected in their meaning, ways to incorporate the uniqueness of place or existing spatial pattern in any distributed models could make model more representative to the watershed conditions and might avoid uncertainty in model prediction (Grayson and Bloschl, 2000). Therefore, effective ways of parameterization of the distributed models by focusing of the uniqueness of the spatial phenomenon could be one of the future directions for the advancement of distributed modeling.

It is often considered that the benchmark level of any distributed models should be the capability to perform at least well than the lumped or black box models (Smith *et al*., 2004). However, in practice, over-parameterization of models is often considered as a root cause of the output uncertainty preventing them from reaching their potential performance level (Grayson and Bloschl, 2000; Perrin *et al.*, 2001; Bashford, *et al.*, 2002). According to the recent alternative blue print of distributed modeling proposed by Beven (2002), any distributed model structure or combination of model parameters should be rejected *a priory* if physically feasible meaning could not be justified. Therefore, major focus of distributed model should be to sort out the practical ways of parameterization so that the numbers of free parameters that require fitting during the calibration are kept to minimum (Refsgaard, 1997). The parameter classes (soil types, vegetations, climatological zones, *etc*.) should be selected in such a way that parameters values associated with the classes could be assigned in objective way, especially, by the measurement on the field. The number of parameters that require calibration should be kept to minimum, and if possible, acceptable range of values of such parameters should be found by measurements. Such approach would result into parsimonious (or less complex) models that could possibly minimize the uncertainty in the estimated output (Wagner *et al*., 2002).

2.6 Remote sensing and GIS application in distributed modeling

Remote sensing and GIS have emerged, in the recent decade, as new techniques to deal with spatial phenomenon. Remote sensing is an indirect method of data acquisition

that relies on response of environmental variables to the incoming electromagnetic radiation (Curran *et al*., 1998). GIS could be defined as a system to store, manage, display, query, and analyze all sorts of spatial information usually coded by the geographical coordinates.

In initial days remote sensing were limited by the use of air photographs, however, after the launching of number of earth observing satellites scope of remote sensing have increased a lot. Description of major applications of remote sensing is out of the scope of this study, so only major applications of remote sensing in hydrology or distributed modeling will be discussed. As explained in last sections that the robustness of distributed model depends on how uniqueness of spatial pattern is incorporated in the distributed model. Recently, number of studies has assessed the possibility of applying remote sensing technique in the distributed modeling (or in hydrology) all of which were mainly focused on minimizing the model uncertainty due to the uniqueness of the places (Schultz, 1988, 1993; Kouwen *et al*., 1993; Bashford *et al.*, 2002; Schmugge *et al*., 2002; Wegehenkel *et al*., 2005). Compared to the range of application of remote sensing in several fields, major application of remote sensing are mainly to derive land cover, vegetation parameters (*e.g.*, *Lai*), precipitation, soil moisture, temperature and snow melting, evapotranspiration *etc* (Houser, *et al.*, 1998; Biftu and Gan, 2001; Andersen *et al*., 2002; Droogers and Kite, 2002; Chen *et al.*, 2005; Zhou *et al.*, 2006). Many of these techniques are still in the process of advancement. In future, it is expectable that remote sensing could be the major source of distributed data for majority of ungauged watersheds (Bloschl, 2005).

Provided huge basin scale data that distributed models need to incorporate, application GIS is indispensable in any distributed models. According to Burrough (1998), at the present context, for the integration of GIS in distributed model number of issues need to be resolved. First, the data in the GIS may not be recorded or stored in the most suitable form for the model. Second, unless someone is efficient in computer programming, designing GIS system to fit the modeling requirement is very time consuming and inefficient. Finally, most visualization methods of GIS do not support interactive space-time presentation of model results. GIS techniques that are specially designed to analyze the hydrological significant watershed variables, such as ArcHydro (Maidment, 2002), or distributed models, such as NSPS modeling (Meiner, 1996), SWIM (Krysanova *et al*.,1998), AGNPS (Young *et al*. 1989) are the increasing examples that are integrated with the GIS.

2.7 Assessment of the model performance

Model calibration (getting values of parameters) and validation (confirming applicability and accuracy) are two crucial steps in modeling that are often used in the assessment of the model. General expectation after successful calibration and validation would be that model could be applied for range of rainfall-runoff conditions in future. However, calibration and validation of distributed models is extremely difficult task because models are highly complex in structure and contain numerous parameters. Because of practical difficulties, only conventional strategies by comparing simulated results with the observed data at key points of the watershed are carried out for both calibration and validation in most of the cases (*e.g.*, Anderton *et al.*, 2002; Vazquez *et al*. 2002; Pebesma *et al.*, 2005). The only difference between calibration and validation lies in the fact that parameter values could be changed to minimize the deviation between estimated and observed data in the former case, while in latter the optimization were not performed. However, in validation, some modification in the model structure could be acceptable provided they were not based on the observed data.

In most of the cases, the judgment of the performance of the model (model structure combination and parameter set) is done using objective function, in combination with visual inspection of the calculated hydrograph. Objective functions aggregate the model residuals, *i.e*., the part of the observed flow not reproduced by the model, as shown by following relation (Wagener *et al*., 2002):

$$e(t|\theta) = y(t) - \hat{y}(t|\theta) \qquad \textbf{2.5}$$

Where, $\hat{y}(t|\theta)$ is the calculated runoff at time step t using parameter set θ, $y(t)$ is the observed runoff at time step t, and $e(t|\theta)$ is the resulting residual at time step t using parameter set θ. The main task is to minimizing the residual, which is mostly done by Simple Least Square (SLS) function,

$$SLS(\theta) = Min\left[\sum_{t=1}^{N} e(t|\theta)^2\right] \qquad \textbf{2.6}$$

Where, N is the number of available observations. Several criteria, such as Nash-Sutcliffe

efficiency, Root mean square error, Bias, Correlation, are routinely used to evaluate the estimation by the model based on the comparison between observed and estimated hydrographs (Wagener *et al*., 2002; Lee and Singh, 2005). The importance of each criterion is mainly based on the distribution and quality of observed data. Especially, confusion arises when values of two parameter sets give similar values of criteria used, but close inspection of the hydrographs could result into completely different pattern of estimated time series (Wagener *et al*., 2002). Therefore, not only the specific criteria but also the hydrograph pattern should be analyzed. In brief, effective calibration and validation of distributed models should consider internal behavior of the models in terms of simulated pattern of state variables and model output (Refsgaard, 1997; Grayson and Bloschl, 2000).

2.8 Effects of the scale on distributed modeling

Among the issues, effects of scale are probably the most widely assessed in distributed modeling (Wood et al., 1988; Kouwen *et al*., 1993; Franchini, *et al.*, 1996). The main reason for that was the dependency of parameters, used in the governing equations, to the scale of the grid. Because of lacking of concrete relationship between the model estimation and scale of the grid, it is often suggested that new model structure should be developed after assessing the scale of data used (Beven, 1995; Bashford, *et al.*, 2002). Apart from assessing the effect of grid sizes on the model estimation, several alternatives were proposed to deal with the issue of scales by using homogenous representative units, such as HRU (hydrologic response units) (Leavesley and Stannard, 1990), REA (representative element areas) (Wood et al., 1988), or GRU (grouped response units) (Kouwen *et al*., 1993), that could be created by overlaying hydrologically significant variables such as land cover, soil, geology, topography etc., by using GIS techniques. However, such approaches were mostly constrained by lacking of effective ways to differentiate the representative area of homogenous representative units.

In general it is perceived that finer scale data would represent more physically based structure, if we assume that the discharge observed at watershed scale were the integrated response from infinitely small units comprising of all significant runoff processes, such as, surface, sub-surface, and underground. For example, some have reported that at finer scale grids the performance of the TOPMODEL was more accurate than at coarser grids (Wolock

and Price, 1994; Bruneau *et al*., 1995). Getting data for the finer grids are still practically difficult therefore implying to depend on the information of coarser grids or point scale measurements. However, a lack of understanding of the nature and extent of grid-scale effects on the properties of watershed limit the use of coarser scale grids in hydrology (Armstrong and Lawrence, 2003). Therefore, the model structure at a given place should be defined beforehand considering available data at different scale. In future, advancement of technology, such as computing capability, GIS and remote sensing might replace the need to model at coarser scale thereby enhancing understanding of effects of scale in distributed modeling processes.

2.9 Summary

Based on above discussions, following points could be emphasized for the current and future researches:

1) Approaches are required to understand the river water quantity and quality, in effective ways, from both modeling and management perspectives. Especially, the river water quality variation caused by the natural influences in the forested watersheds of Japan need to be emphasized to assess the long-term sustainability;

2) Need of simple and less parameter intensive parsimonious distributed models that could overcome the estimation uncertainty due to over parameterization;

3) Utilization of techniques, such as remote sensing and GIS, to generate distributed information and properly manage the spatial information, could increase the scope of distributed models in future;

4) Issues of scale are certain to increase the estimation uncertainty by the distributed models. For the present context, it is therefore quite essential to assess the effect of scale on the performance of distributed model in addition to the conventional calibration and validation;

CHAPTER III

DESCRIPTION OF STUDY AREA AND DATABASE CONSTRUCTION

3.1 Introduction

Data of a watershed are usually spatial in nature. Spatial data such as elevation, land use, soil, and geology could be linked directly or indirectly with the hydrological and biogeochemical processes occurring inside the watershed. Especially, in distributed models, major input parameters are based on the spatial information of the watershed such as rainfall, land use, soils, topography *etc*. Therefore, construction of spatial datasets and their management are fundamental to accomplish overall goal of this study. Based on the characteristics of constructed database, appropriate analysis techniques could be adopted, thereby making the latter tasks more effective.

This chapter will firstly give a brief overview of the study area and its important characteristics. However, major focus will be on the preprocessing of spatial databases.

3.2 Methodology

3.2.1 Study area and its general characteristics

Kamo watershed was selected as a test study site, which was shown in the **Fig. 3.1**. The watershed is located at the north of Kyoto city (latitude $35^0$01'-$35^0$11'N; longitude $135^0$42'-$135^0$52'E) and its major river system is commonly know as Kamo river, which flows towards the south. It has an area of 142.2 km^2 and most its characteristics are typical to the hilly watersheds found in Japan. The watershed was dominantly covered by forest areas (> 85 %) most of which were on hilly parts. Urban areas (~ 11 %) were the second largest land cover type. Urban areas were usually connected by centralized sewer system, comprising of both combined and separated sewers as shown in **Appendix A(5)**. All of the water and loadings from the combined sewer were transported outside the basin so they were excluded from the analysis, *i.e.*, Fig. 3.1 represented the area after excluding the com-

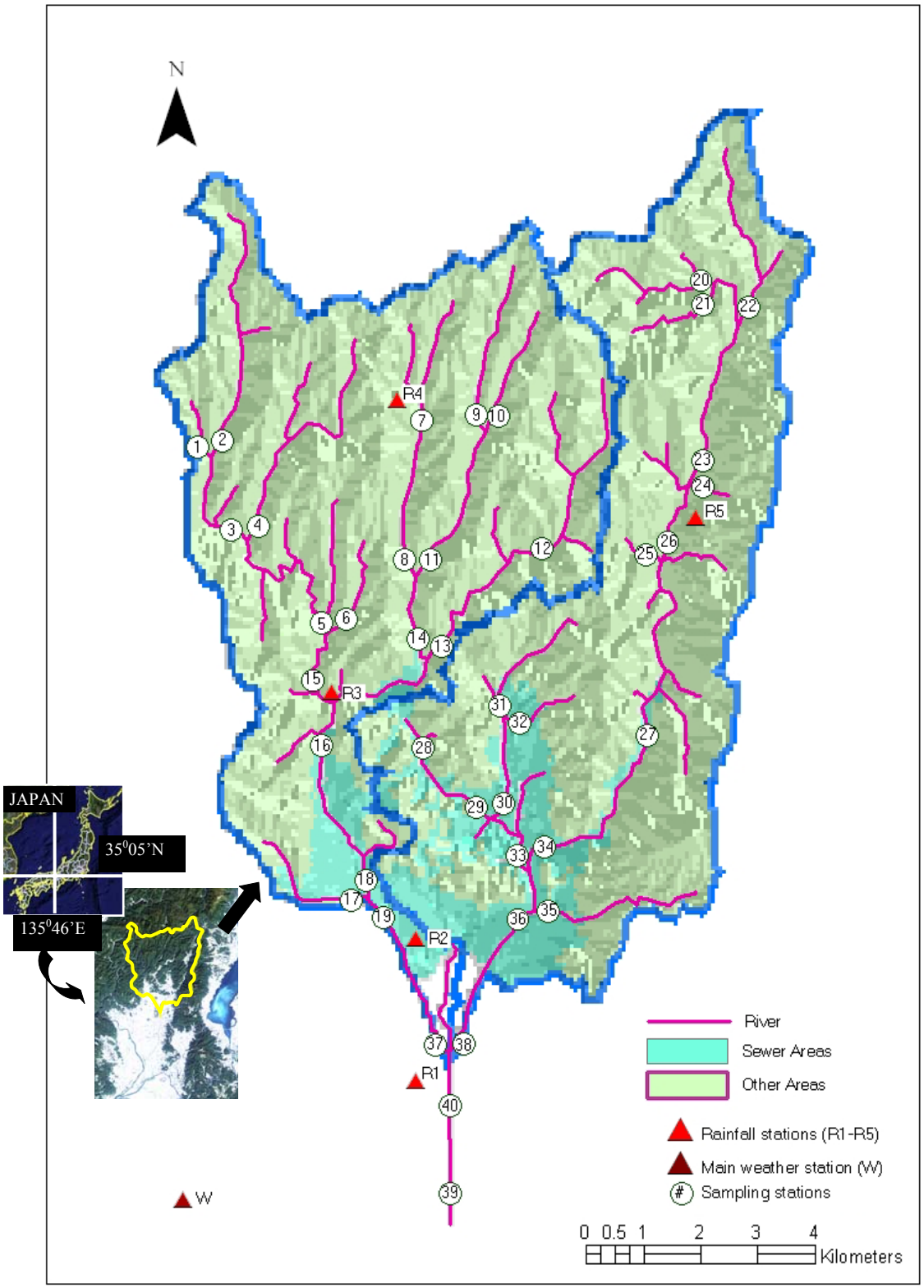


Figure 3.1 Map of the study area with relevant attributes

bined sewers. For the separate sewer areas, only wastewater (especially domestic loadings) were transmitted outside of the watershed, but rainwater was conveyed to the main river through drainage pipes or channels. Therefore, river was not affected directly from the domestic loadings. Some parts were also used for agriculture (~ 3%), in which paddy was the most prominent type. Soil type of the study area was acid brown forest soils (Dystric or Utric cambisols, FAO soil classification) (Nagatsuka and Okazaki, 2005), which was also the major soil type found all over the forested areas of Japan.

Figure 3.2 shows the trend of river flow rates and major water quality indices (SS, pH, COD_{Mn}, TN, NO_3^-−N, and TP) for the period of nearly 11 years (1988-1999) observed at different stations, in which each plot represented one month. The data was obtained from the local office of Kyoto municipality, which had been continuously monitoring the river water for long period. At the outlet station (St. 39), flow rate showed some variation, but majority of water quality indices did not showed much variation there. For other stations, the profiles were quite similar for both flow rates and water quality indices and they did not showed remarkable change in all periods. Especially, pH was nearly neutral for all periods, which might indicate the unimpaired river water quality in the whole periods. However, indices such as COD_{Mn}, TN, and TP showed decreasing trend since the initial period towards 1999 at the St.33, whose upstream areas were mainly urban parts. In addition to that, all three indices (COD_{Mn}, TN, and TP) were also inter-correlated ($r > 0.8$) among each other at St.33, which could be due to the possible domestic loadings. One of the possible interpretations could be due to the improvement in the sewer facility in the watershed that caused gradual decrease in the concentration of these water quality indices. However, relatively similar concentration at the outlet in all periods could mean that those effects were usually normalized.

3.2.2 Construction of spatial database

Spatial data were needed in this study for both water quality assessment and distributed modeling. Major spatial data used in this study were land use (or land cover), geology (including soil profile depth), population, elevation, and sewer maps, rainfall and other daily weather variables. One of the essential requirements, in our study, was to convert all watershed attributes into grid (raster) format, so that spatial information could be assigned in the distributed model. The basic grid size was nearly 50m (46.21 × 56.91m),

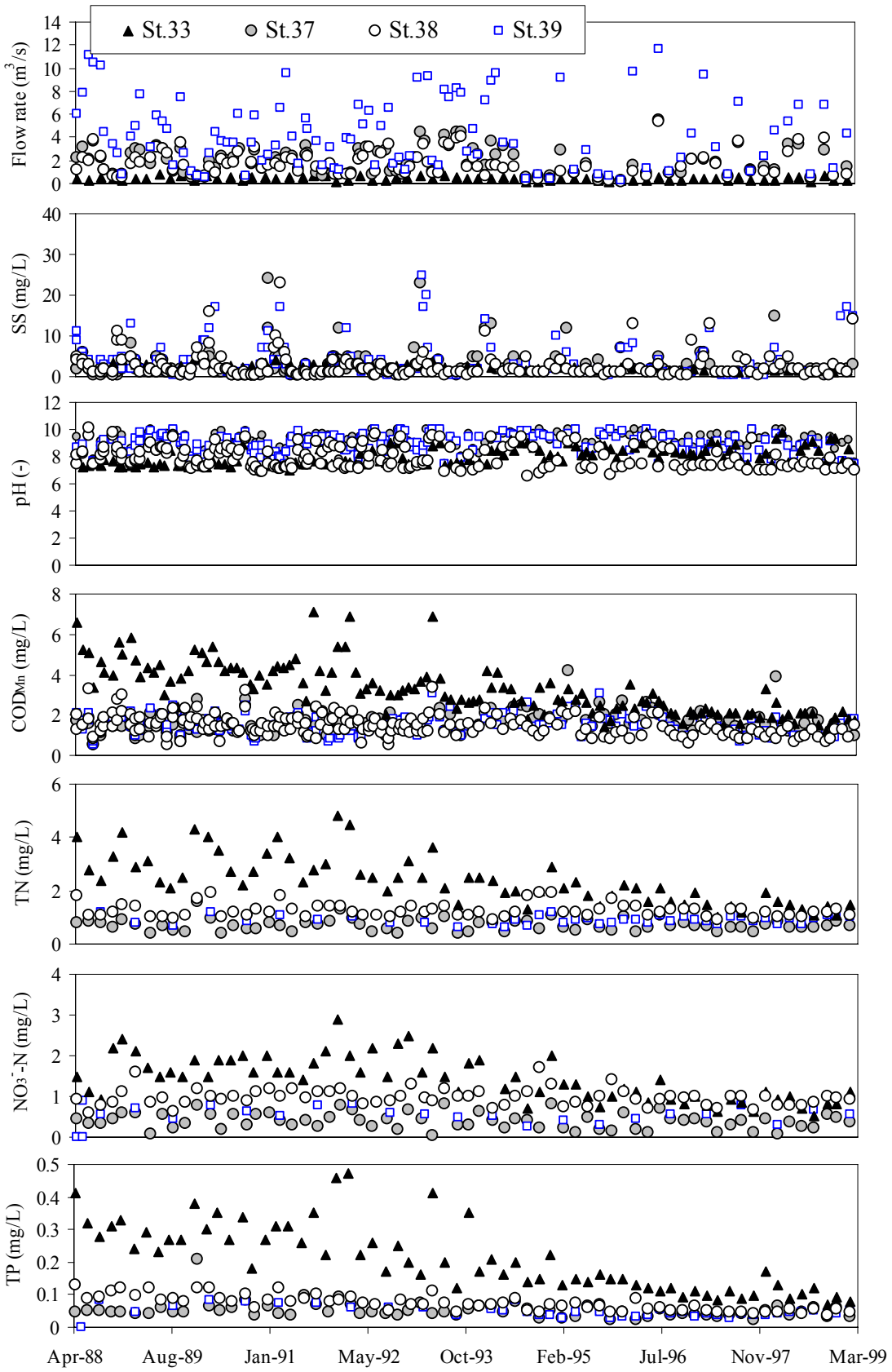


Figure 3.2 Trends of flow rate and major water quality indices at four stations for 1988-1999

which was the original scale of the of digital elevation model (DEM) (GSI, 1997). Hereafter 50m will be used to represent the size of base grid for the sake of convenience, but real dimensions will be used for all processing and calculations. All of the processing were done mainly in *MS-Excel* and accompanying *VBA* (*Microsoft Inc.*), but some were also done by using *EVNI* 4.2 (*Research Inc.*) or *ArcGIS 8.3*(*ESRI Inc.*) and are specified wherever used.

Meteorological data (rainfall, air temperature, relative humidity, wind speed, and sunshine duration) observed at the main station (W) located outside of the watershed (~ 6 km southwest) was downloaded from the Japan meteorological business support service center web page (Fig. 3.1) (JMC, 2007). In which, both daily and hourly rainfall data were collected, but data of other meteorological variables were collected on daily basis. In addition to that, daily rainfall data were collected from the other five stations (R1~R5) inside the watershed (Fig 3.1). Each grid was assigned its rainfall data from the nearest rainfall station. The distance was calculated by using following equation:

$$d_{R,m} = \left\{ \left((R_i - m_i) \times 46.21 \right)^2 + \left((R_j - m_j) \times 56.91 \right)^2 \right\}^{0.5} \qquad \textbf{3.1}$$

Where, $d_{R,m}$ was the distance between R^{th} rainfall station and m^{th} grid, i was the row number and j was the column number in a worksheet of *MS-Excel*.

Paper maps of land use (1: 25,000 scale) (GSI, 1999) and geology (1: 50,000 scale) (Kyoto Prefecture, 1982) were available, which were later on digitized into digital grid format. For the digitization, paper maps were at first divided into basic grid size considering the scale of the maps, which were then coded by unique integers to represent specific land cover or geology category in the *MS-Excel* worksheets.

In addition, satellite images were also processed in order to get land cover map and vegetation variables (details in Chapter VI). Four nearly cloud free images, of the Landsat 7 (ETM+), taken at different months were used, which is shown in **Table 3.1**. Image of 8th December had some missing areas (10 %) but it was better quality, so missing areas were merged from the image of 15th December. Further details of all four images are depicted in the **Appendix A (1-4)**. Landsat ETM+ has nine bands, whose relevant properties and major application fields are summarized in the **Table 3.2**. Among nine bands, six are thematic bands (*TM*1 ~ *TM*5, *TM*7) that have same spatial resolution. Each thematic band

Table 3.1 Landsat ETM+ images used in this study

Acquisition date	Row	Path	Correction level
15-October-2001	36	110	1G
15-December-2000	36	110	1G
05-May-2000	36	110	1G
08-December-2000	36	109	1G

has some specific application areas, but their combined analysis could detect many other environmental variables. In our study, we mainly focused on applying the thematic bands, because major parameters derived from the analysis of images were related with land cover or vegetations. Two are thermal bands (*TM*6.1 and 6.2) that have spatial resolution coarser

Table 3.2 Bands of Landsat ETM+, their relevant properties, and major application areas

Band Type	ID	Common name	Wavelength (μm)	Grid resolution (m)	Major application areas
Thematic	*TM* 1	Blue	0.45-0.52	28.5	Water bodies, soil/vegetation discrimination, forest mapping
	TM 2	Green	0.53-0.61	28.5	Green color reflectance
	TM 3	Red	0.63-0.69	28.5	Chlorophyll absorption
	TM 4	Near infrared (IR)	0.78-0.90	28.5	Vegetation, water bodies, soil moisture
	TM 5	Mid-IR	1.55-1.75	28.5	Moisture in vegetation, soil etc
	TM 7	Far-IR	2.08-2.35	28.5	Mineral, rocks, vegetaion moisture
Thermal	*TM* 61 (low gain)	Thermal-IR	10.4	57	Thermal mapping, moisture stress of vegetaion, soil moisture
	TM 62 (high gain)	Thermal-IR	12.5	57	
Panchromatic	*TM* 8	__	0.5-0.9	14.25	Pansharpening (combine low and high spatial resolution bands)

Source: Lillesand and Kiefer (2000)

than thematic bands. These two bands were mainly used in the analysis of thermal properties of the features, so they were not considered in this study. *TM*8 band has the highest spatial resolution among all bands, which has its main application to combine the lower resolution bands so that the resulting image would have higher resolution without significantly losing the spectral properties. *TM*8 was mainly used to draw the vector river line in this study (details in Chapter VI). Before the analysis of images, all available bands were converted to at-sensor reflectance by using Landsat-7 Science Data User handbook

(http://landsathandbook.gsfc.nasa.gov/handbook/handbook). Reflectance was used instead of the radiance (mW/cm^2.sr.μm) because it was a relative measure and usually preferred in the analysis of multiple date images.

Population data was collected from Kyoto city office but they were available for each sub-wards (smallest administrative area) rather than in grid format. Since urban and agricultural areas usually represented the places having higher population density, so it was possible that areas in those grids could be related with them. Then a multiple regression model, without intercept, was introduced to determine the population for each grid. Following equation depicts the regression model used in this study.

$$Y_k = p_1 x_{1k} + p_2 x_{2k} + \ldots + p_i x_{ik} + \ldots + p_n x_{nk} \qquad \textbf{3.2}$$

Where, Y_k was the total population (capita) of k^{th} sub-ward in the year 1999, p_i was the coefficient or slope (capita/km^2) for i^{th} land use category, and x_{ik} was the area (km^2) of i^{th} land use category of k^{th} sub-ward, and n was the total number of land use categories.

Sewer map was provided by the Kyoto city office, which was later on digitized into grid format (similar method with land use or geology), which is shown in **Appendix A (5)**. Elevation map was available in digital grid format (DEM), so it was directly processed to get topographic variables, such as drainage direction, watershed boundary, and sub-basins related with each station. Among them, drainage direction was the most important topographic variable that was indispensable in the distributed modeling because it was the only basis to route the water and pollutants from upstream grids to the downstream grids.

Drainage direction was determined by the D8 method after preliminary processing (*e.g.*, removal of pits or adjustment of plural directions). However, due to the differences between the scale of DEM and real topography, drainage direction determined by this method sometimes could not coincide with the drainage features inside the watershed. Similarly, in our study, it was essential that all the sewer cells remain interconnected to each other. Therefore, we additionally create a few more criteria other than D8 method:

a) For a normal grid water flowed according to D8 method;

b) In case of sewer grid, the downstream steepest gird was either sewer or river grid, but not the normal grid;

c) All the water and pollutants from normal or sewer grids finally terminated into the river grid;

It was also important to separate the contributing sub-basin area of 39 sampling stations. For that purpose, number of grids contributing to each sampling station was separated by utilizing the drainage direction map. The flow chart of drainage direction determination and sub-basin separation are shown in **Fig. 3.3,** however further details are explained in Chapter VI, including the derivation of slope and topographic index of TOPMODEL.

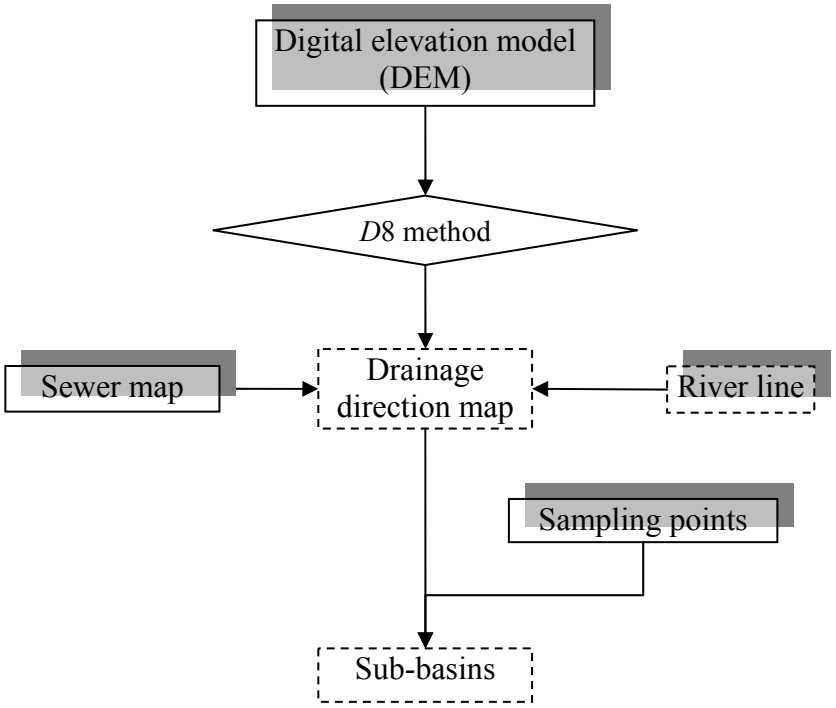


Figure 3.3 Determination of drainage direction and separation of sub-basins

3.3 Results

3.3.1 Rainfall and climate

Figure 3.4 shows the map of study areas classified with respect to the nearest rainfall station. Rainfall stations R2-R5 shared the maximum basin area, while R1 just shared a little area. All rainfall stations showed similar rainfall pattern with main station (W) showing correlation (*r*: 0.86 -0.99), therefore assignment of the rainfall to each grid based on the nearest rainfall was assumed representative. **Figure 3.5** shows the long term distribution of rainfall in for the period of 125 years (1881-2006). It could be observed that annual distribution pattern followed nearly sinusoidal pattern, in which cyclic rainfall patte-

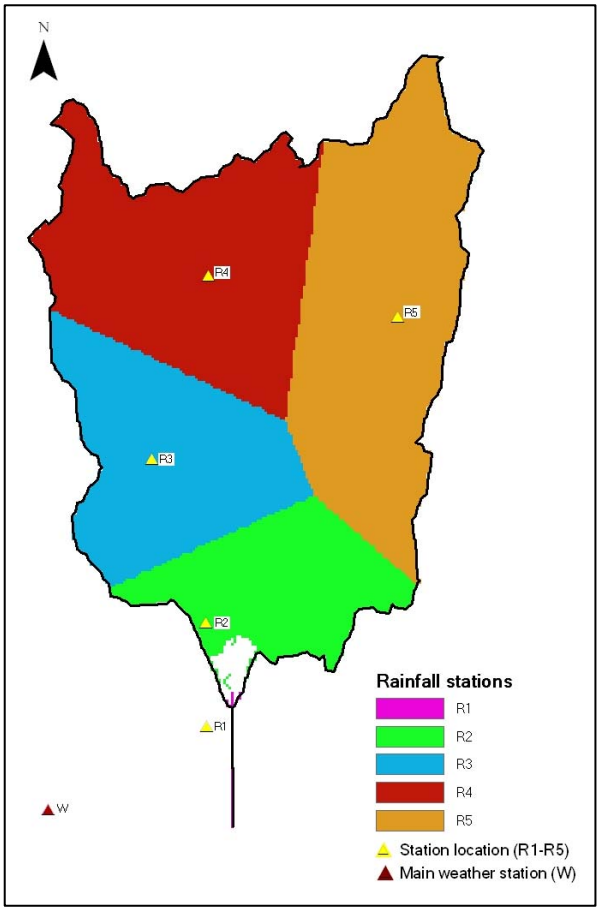


Figure 3.4 Watershed area classified with respect to the nearest rainfall stations

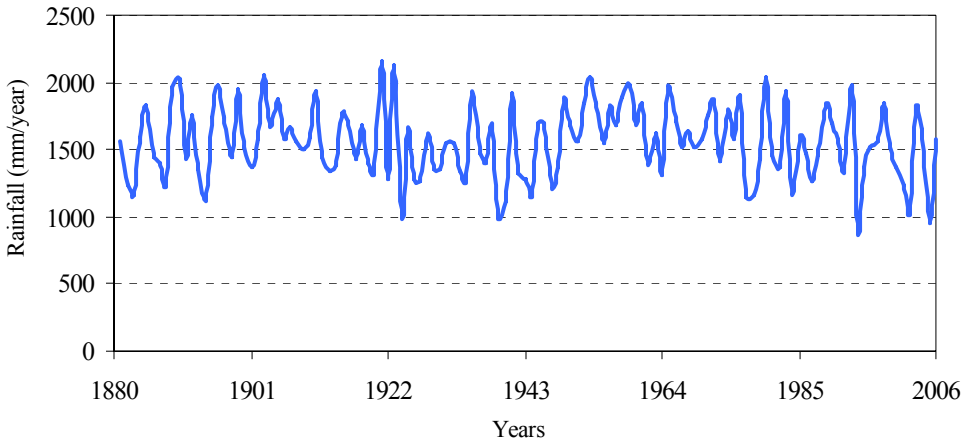


Figure 3.5 Annual rainfall distribution pattern in the study area for the period of 125 years(1881-2006)

rn of nearly 4 - 8 years was observed. In the whole period, the driest year received 880 mm/year of rainfall while in the wettest year the maximum rainfall of 2160 mm/year was

observed. In terms of monthly rainfall distribution, nearly five months (November - March) were the driest months in which the minimum rainfall of 2.5 mm/month was observed, where as in the remaining wetter months the maximum rainfall as high as 626.9 mm/month was observed in the whole period. The historic rainfall trend showed that the watershed experienced both drier and wetter periods in nearly cyclic yearly pattern. The pattern indicated that the future watershed management should focus on both the extreme and driest rainfall runoff condition, which could be more severe in future due to the increasing global warming. In addition to the historic data, detail analysis of rainfall pattern was also required. **Figure 3.6** shows the daily rainfall and temperature (maximum, average, minimum) distribution observed at the main weather station (W) for one year (2003). The rainfall seemed to appear frequently in the area. The occurrence of the rainfall was 141 days/year, which was equivalent to nearly five months of continuous rainfall. The rainfall intensity was also higher in which maximum rainfall intensity of 115 mm/day (31 mm/hr) and average rainfall intensity 13.17 mm/day (3.1 mm/hr) was observed during the whole year. The average duration of the rainfall was only 6.12 hours, but the maximum duration of continuous rainfall even reached up to 32 hours. Especially, frequent rainfall may keep the watershed moist for most of the times while intense and short duration rainfall events could result into flashy flow regimes. Both daily and seasonal variation in temperature could be seen on the area, which could mean that weather condition could be quite different even between consecutive days. In addition to that, wide seasonal variation in temperature, that reached 35^0C in the summer period and below 0^0C in winter period, could indicate that watershed was also influenced by significant change in the climatic conditions. Frequent

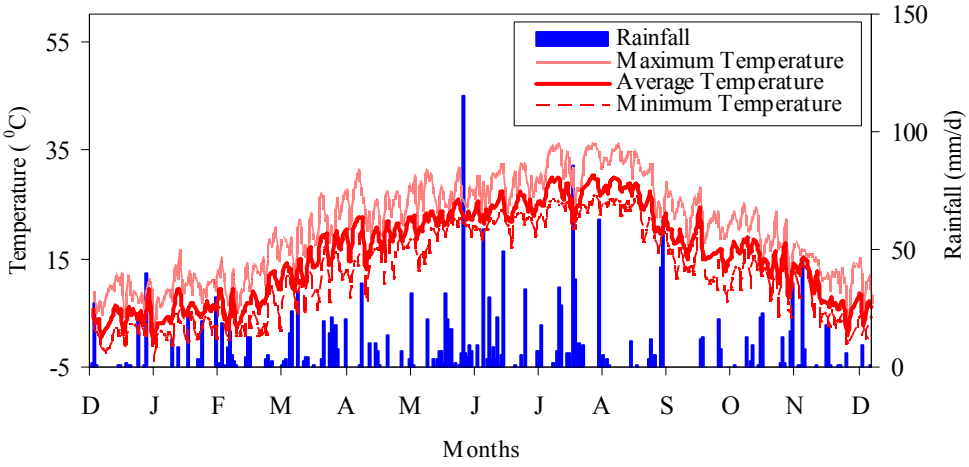


Figure 3.6 Daily rainfall and temperature distribution in the study area for the year 2003

rainfall and wider variation in the climatic condition could be quite important for the biochemical processes, such as weathering of rocks and mineralization inside the watershed. In addition, differences in the climatic conditions could significantly affect the evapotranspiration rate, which could have significant influence on the water balance of the whole watershed.

Table 3.3 Land use attributes and population proportion in the contributing area of 39 sampling stations

Stations	Sub-basin, km^2	Land-use			Population			
		Forest, %	Agriculture, %	Urban, %	Total, %	density, ca/km^2	No sewer, %	Sewer served, %
1	1.50	100	0	0	0(0)	0(0)	0(0)	0(0)
2	6.20	100	0	0	0(0)	0(0)	0(0)	0(0)
3	10.40	98	0.3	1.7	0.17(100)	18(71)	100(100)	0(0)
4	7.40	100	0	0	0.05(100)	7(7)	100(100)	0(0)
5	22.90	99	0.18	0.82	0.3(27)	14(17)	100(27)	0(0)
6	1.70	100	0	0	0(0)	0(0)	0(0)	0(0)
7	4.20	100	0	0	0(0)	0(0)	0(0)	0(0)
8	7.20	99	0	1	0.1(100)	15(35)	100(100)	0(0)
9	3.20	100	0	0	0(0)	0(0)	0(0)	0(0)
10	2.30	100	0	0	0(0)	0(0)	0(0)	0(0)
11	8.80	98	0.12	1.88	0.51(100)	64(172)	100(100)	0(0)
12	7.00	99	0.9	0.1	0.14(100)	22(22)	100(100)	0(0)
13	12.90	93	4.49	2.51	0.77(82)	65(117)	100(82)	0(0)
14	17.30	98	0.06	1.94	0.81(25)	52(188)	100(25)	0(0)
15	29.40	99	0.21	0.79	0.48(38)	18(42)	100(38)	0(0)
16	64.80	96	1.33	2.67	5.14(60)	87(651)	46(6)	54(100)
17	2.20	62	0	38	6.4(100)	3222(3222)	1(1)	98(100)
18	69.80	93	2.06	4.94	17.46(71)	275(2709)	16(3)	84(81)
19	72.80	92	1.98	6.02	28.16(15)	426(5838)	11(0)	89(14)
20	4.70	100	0	0	0(0)	0(0)	0(0)	0(0)
21	2.20	100	0	0	0(0)	0(0)	0(0)	0(0)
22	4.60	99	0.23	0.77	0(100)	0(0)	100(100)	0(0)
23	18.10	94	1.62	4.38	0.3(99)	18(49)	100(99)	0(0)
24	1.40	88	3.85	8.15	0.23(100)	188(188)	100(100)	0(0)
25	1.10	89	5.88	5.12	0.16(100)	166(166)	100(100)	0(0)
26	27.40	90	4.68	5.32	2(73)	81(205)	100(73)	0(0)
27	35.50	91	4.24	4.76	3(24)	88(106)	84(8)	16(100)
28	2.20	82	5.74	12.26	2(100)	843(843)	8(8)	92(100)
29	5.00	69	8.54	22.46	7(76)	1504(2014)	13(11)	87(74)
30	8.10	77	6.73	16.27	9(22)	1201(4423)	3(0)	97(23)
31	6.10	83	5.33	11.67	5(100)	889(889)	5(5)	95(100)
32	1.50	76	2.08	21.92	2(100)	1410(1410)	0(0)	100(100)
33	16.20	71	6.89	22.11	22(28)	1492(2261)	5(0)	95(29)
34	42.50	89	3.86	7.14	11(73)	275(1217)	28(6)	72(94)
35	2.30	92	0.91	7.09	1(100)	660(660)	20(20)	80(100)
36	65.40	81	4.6	14.4	53(35)	886(4733)	9(1)	91(38)
37	75.40	89	2.05	8.95	41(31)	594(5330)	7(0)	92(33)
38	66.80	80	4.53	15.47	59(11)	977(5027)	8(0)	91(9)
39	142.50	85	3.21	11.79	100(0)	772(0)	8(0)	92(0)

NB:Figures in parenthesis were obtained after excluding upstream sub-basin contributions

3.3.2 Distribution of spatial data

The watershed had hilly topography with elevation range (maximum-median-minimum) of 890-363-46 (m). **Figure 3.7a** shows the DEM of the study area, in which most of the upstream hilly areas had elevation greater than 300m. The areas were relatively flat with the elevation range of 40-300 m near the main river or near the main outlet. **Figure 3.7b** showed the map of sub-basins associated with each station. The relevant attributes of each sub-basin are given in the **Table 3.3**. Land use also followed the topography, in which majority of sub-basin were covered by the forests. Especially, the stations at the upstream usually had more than 90% area covered by the forests (**Figure 3.7c**). Urban and agricultural areas were mostly concentrated towards the downstream areas that were relatively flat. Urban areas mainly included the big non-residential buildings, big residential apartments, crowded residential areas, and non-crowded residential areas. **Table 3.4** showed the results of regression analysis, in which five main land use categories, without forest areas, gave a good relationship with high R^2 of 0.97 and *p-value* of *F-test* of 1.5×10^{-18}, indicating possibility of significant reliability of the model to estimate the population for each grid type.

Table 3.4 Summary of regression analysis to determine the raster map of population distribution

	Land use	Coefficinets (a_i)	*p-value*
	Agriculture (x_1)	19.9	0.84
Urban	Buildings (non-residental) (x_2)	34.7	0.37
	Buildings (residental) (x_3)	359.4	< 0.001
	Residental houses (crowded) (x_4)	131.9	< 0.001
	Residental houses (not crowded) (x_5)	71.9	< 0.001

At first, the regression equation was applied to determine the population number in all sub-wards, in which correction factors were applied to the equation to conserve the population number in all sub-wards. By applying the coefficients of regression model, total 110,070 capita of population was determined in the whole watershed. The distribution of the population is depicted in **Fig. 3.8a**. Especially, the urban areas had a very high population density (4423 capita/km^2 at St. 30) in comparison with the population density of the watershed that was only 772 capita/km^2. It was due to the values of coefficients of regression equation (capita/km^2), in which residential category had very high value. Of total population, nearly 92% were connected to the sewer service facility, but 8% of the

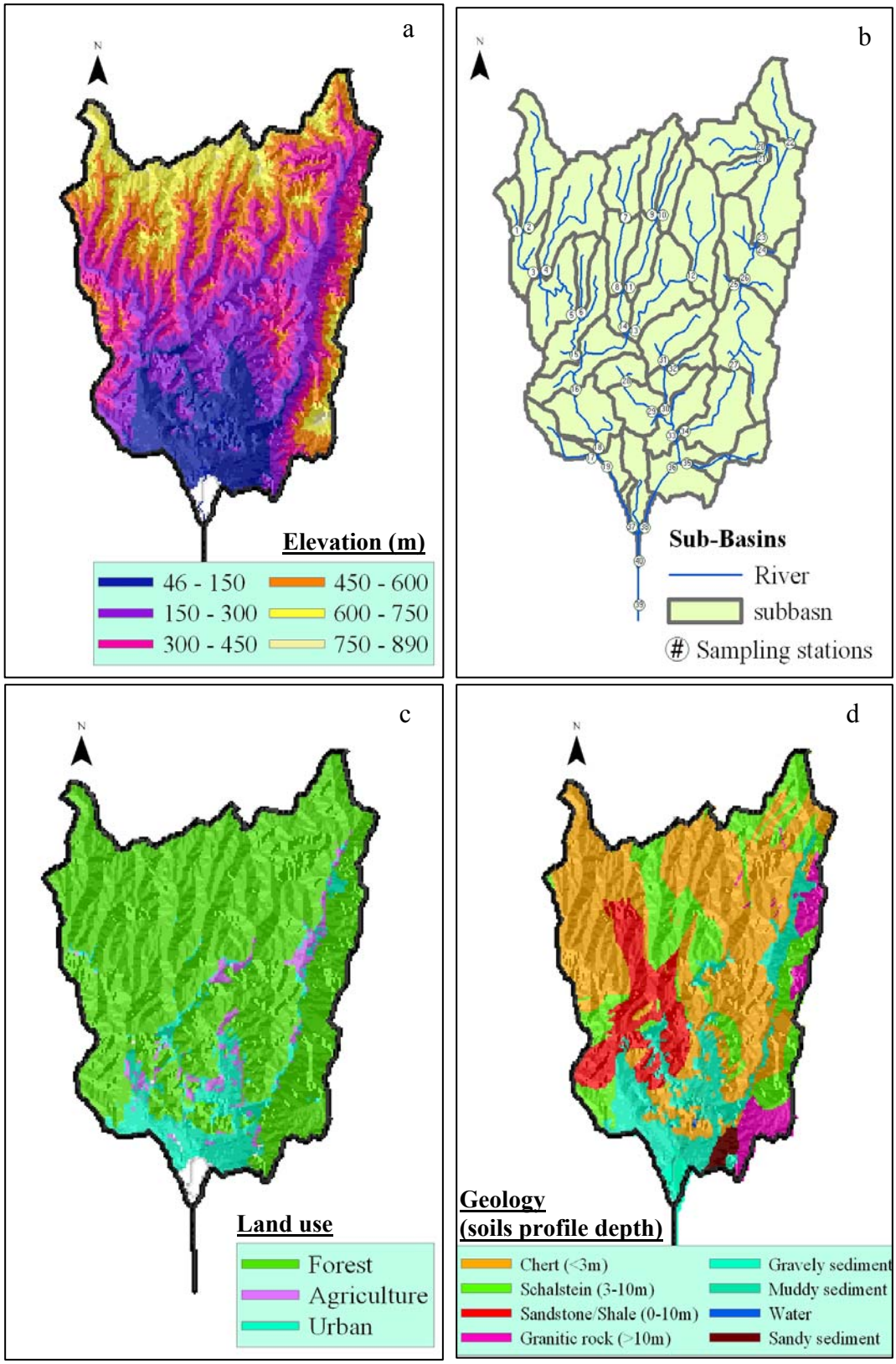


Figure 3.7 Different raster maps used in this study [a) Elevation; b)Contributing sub-basins; c)Land use; d) Geology with depth of soil profile]

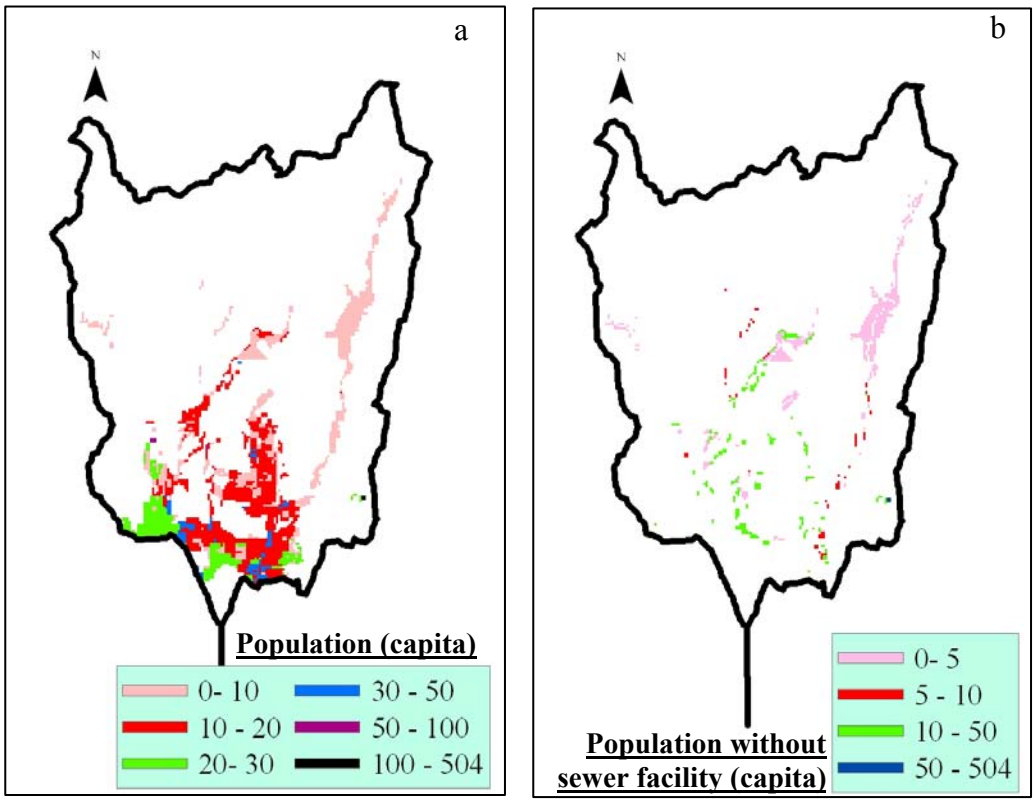


Figure 3.8 Population maps [a) Total population; b) Population without sewer facility]

population was not connected to the sewer facilities. Therefore, only 8% of the population, without sewer facility, was considered to assign the loadings for the respective grid during the modeling process. Those areas without sewer facilities were mainly concentrated at the upstream parts, whose population distribution is shown in **Fig. 3.8b**.

The geology was mainly composed of chert, schalstein, sandstone, shale, granite, and others (greenstone, mudstone, sandy sediments, muddy sediments, gravelly sediments *etc*) which is shown in **Fig. 3.5d**. All these rocks have the highest silica content (50~90%), as compared with other minerals, such as Ca, Na, K, Al, Mg, Mn and Ti, each of which is usually less than 10% of the total composition. The soil profile depth (weathered section of geology) could be classified into three groups, less than 3m, 3-10m and more than 10m, of which majority of soil profile (>70%) were shallow (< 10m). Generally, areas (46%) having chert type geology had less than 3m of soil profile above it. While majority of areas (31%) having sandstone, shale, and schalstein had 3-10m soil profile depth, while granite (5%) even had more than 10m soil profile depth. Depth of other areas having sediments (gravel, sand, mud) were not available, so they were assumed >10m.

3.4 Summary

Results of any analysis depend on the quality of database used. It was therefore necessary that all of the database were carefully prepared and managed before the analysis. Besides, management of database could be considered as preliminary assessment because it gives a general idea for the latter analysis. In this chapter relevant databases (spatial, meteorological, population) were prepared. All of the data were then converted into grid format, which was an essential requirement for the distributed model.

From the overview of the spatial database of the study area, the characteristics seem to be similar to the other watersheds found in the Japan, especially:

-Frequent and intense rainfall events, and wide variation in day-to-day as well as seasonal climatic conditions;

- Hilly topography with relatively shallow soil profile having acid brown forest soil type

- Majority of hilly parts were covered by forests, where as the plain areas houses almost all of the population that were connected with the sewer facilities;

CHAPTER IV

WATER QUANTITY AND QUALITY SURVEYS AND THEIR SPATIAL AND TEMPORAL VARIATION ANALYSIS

4.1 Introduction

River water quantity and quality generally varies at different spatial and temporal conditions. Water quantity and quality data are usually collected and later on analyzed to understand their variations at different spatial and temporal conditions. Analysis of the collected data could also provide a basis whether the collected data could be effectively utilized during the construction and verification of the distributed model.

Important aspects of data analysis were to understand the impacts of natural as well as anthropogenic impacts on water quantity and quality. In case of Japan, considering dominant forest cover, steep topography, frequent rainfall, and acid rain, it was very essential to understand the variation pattern of river water quantity and quality both from future perspective and to formulate hypothesis during the modeling.

In this chapter, we mainly applied two multivariate analysis techniques, namely principle component analysis (PCA) and cluster analysis (CA) to analyze the multidimensional river water quality data. Following are the specific points emphasized in this chapter: 1) collection of water quantity and quality data through several pre-planned surveys; 2) consideration of both spatial and temporal water quantity and quality variation; and 3) the role of forest areas on the river water quantity and quality variation.

4.2 Methodology

4.2.1 Collection of water quantity and quality data

Several water quantity and quality surveys were conducted from 2000 to 2006, which are summarized in **Table 4.1**. Surveys could be divided into two types:

1) Spatial surveys

2) Temporal surveys

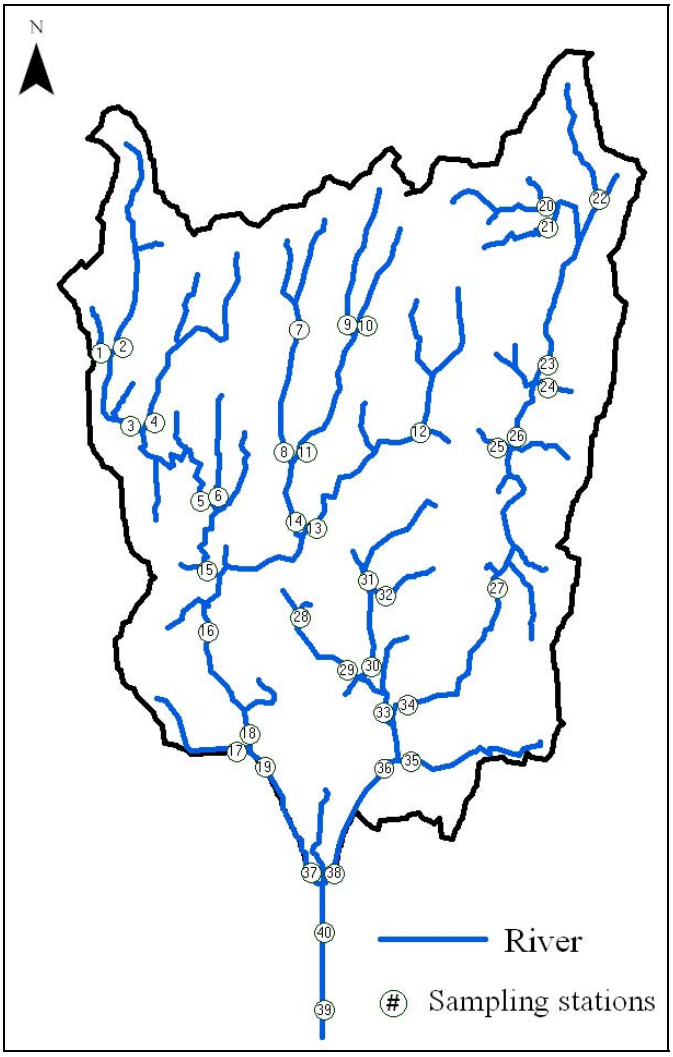


Figure 4.1 Map of sampling stations across the main river network

In spatial surveys, 39 sampling stations (St.1 ~ 39) were established at main river sections across the watershed, which are shown in the **Fig. 4.1**. Altogether six spatial surveys were conducted, and all of them were usually started at 6:00 hours in the morning and completed by 13:00 hours. It was also ensured that the watershed did not receive direct influence of rainfall prior to the sampling. Therefore, the survey was also named as fine weather simultaneous survey. The temporal surveys were carried out at different times of the year at the two outlet stations (St. 37 and St. 38). The surveys could be classified on the basis of frequency of sampling. Samplings were conducted ranging from few hours, alternate days, every 10 days, and monthly interval. Sampling at few hours interval (~4 hours) was conducted especially during the storm events when the rainfall intensity was relatively higher so it was named as storm event survey. Other temporal surveys were

Table 4.1 Details of different types of water quality surveys data used in this study

Surveys	Periods	Frequency	Stations	Total data (n)
Temporal:	'02/09/23-'02/12/12	every 2 days	St. 37 & 38	372
Regular	'02/12/12-'03/12/10	every 10 days		
	'03/12/10-'06/10/10	monthly		
Storm events	'02/10/19,'02/11/07	6 times (~every 4 hrs)		
	'03/07/20,'03/07/22			
	'03/11/05,'03/11/28			
Spatial:	'00/11/24,'01/10/24	6 times	St. 1~39	234
Simultaneous	'02/12/02,'04/05/26			
	'04/10/16,'04/12/02			

Table 4.2 Major water quality indices (WQIs) and their analysis methods

Water quality parameters	Unit	Methods and instruments
River flow rate	(m^3/s)	Propeler type flow meter, DENTAN Elec.
pH	(-)	pH meter
Dissolved Oxygen (DO)	(mg/L)	DO meter
Chemical Oxygen Demand (T-COD_{Mn},D-COD_{Mn})	(mg/L)	Titration method
Dissolved Organic carbon (DOC)	(mg/L)	TOC-5000A analyzer
Inorganic carbon (IC)	(mg/L)	TOC-5000A analyzer
Suspended solid (SS)	(mg/L)	Oven dried at 105 ^{0}C for 2 hrs
Volatile suspended solid (VSS)	(mg/L)	Oven dried at 550 ^{0}C for 5 hrs
Nitrogen (TN,DN)	(mg/L)	BRAN+LUEBBE Auto analyzer
Ammonium Nitrogen (NH_4^+-N)	(mg/L)	BRAN+LUEBBE Auto analyzer
Nitrate Nitrogen (NO_3^--N)	(mg/L)	BRAN+LUEBBE Auto analyzer
Nitrite Nitrogen (NO_2^--N)	(mg/L)	BRAN+LUEBBE Auto analyzer
Phosphorous (TP,DP)	(mg/L)	BRAN+LUEBBE Auto analyzer
Chloride (Cl^-)	(mg/L)	Ion chromatography, DIONEX
Sulfate (SO_4^{2-})	(mg/L)	Ion chromatography, DIONEX
SiO_2-Si	(mg/L)	Molybdate blue method
Sodium (Na)	(mg/L)	ICP/MS
Potassium (K)	(mg/L)	ICP/MS
Magnesium (Mg)	(mg/L)	ICP/MS
Calcium (Ca)	(mg/L)	ICP/MS
Barium (Ba)	(mg/L)	ICP/MS
Strontium (Sr)	(mg/L)	ICP/MS
Iron (Fe)	(mg/L)	ICP/MS
Aluminium (Al)	(mg/L)	ICP/MS
Manganese (Mn)	(mg/L)	ICP/MS

conducted irrespective of weather conditions. Except storm event survey, the time for all temporal surveys was set at 8:00 hours in the morning. In both spatial and temporal surveys, each observation was monitored for nearly 28 water quality indices (WQIs) including river flow rate. The measured WQIs including the methods of analysis are shown in **Table 4.2**. Besides our planned surveys, daily river water level data, observed daily at 500m downstream (St. 40) of the St. 37 (or St.38), was also collected from local government office, Kyoto municipality. The river water level was mainly utilized to assess the river runoff pattern and to compare with the model estimation in Chapter VII.

4.2.2 Analysis of water quantity and quality data

Collected water quality and quantity data were of less meaning unless their effective analysis was performed. Initial attempts of this study were focused on getting an overview of general trend of rainfall-runoff and the concentration differences of WQIs at different stations as well as during different periods. Although simple analysis was quite important to understand the general trend of both water quantity and quality data distribution, they could not provide meaningful conclusions, especially, related with their inter-relationships and variation pattern at different spatial and temporal context. Therefore, all the data should be analyzed together to understand the covariation patterns.

For that purpose, two databases were prepared on the basis of main survey types. All data collected during spatial surveys were arranged to construct the spatial dataset and similarly that of temporal surveys were arranged to construct the temporal dataset. The flow rate (m^3/s) observed at different stations was affected by their sub-basin areas, so they were normalized by the contributing sub-basin area to get the specific flow rate (SFR) (mm/d). Initially, general overview and correlation pattern of all WQIs were presented and the main similarities or differences were pointed out. Then the spatial dataset was applied to three main analysis, namely, two ways layout analysis of variance (ANOVA), cluster analysis (CA), and principle component analysis (PCA) to examine the spatial variation pattern. After that, temporal dataset was applied to PCA and the results were assessed to understand the temporal variation pattern by using scattering diagrams. Both CA and PCA were applied on the standardized data of each WQIs, *i.e.*, average = 0 and standard deviation = 1.

4.3 Results and discussion

4.3.1 Rainfall-runoff pattern

Rainfall-runoff pattern and river water quality are closely linked to each other. Rainfall-runoff processes of the study area were required both during the assessment of river water quality and development of the model, so their preliminary analysis was quite important. **Figure 4.2** shows the average flow rate at 39 sampling stations. The flow rate was shown from upstream to downstream, in which increasing flow rate condition with the increasing contributing areas were obvious, whereas the stations showing smaller flow rates at the downstream were the tributaries. Since all surveys were conducted during fine weather conditions, the observed flow rate could be assumed as base flow condition. At the final outlet station (St. 39), the average flow rate was nearly 3.5 m^3/s. Similarly, at St. 37 or St. 38, the flow rates were generally below 2 m^3/s. Therefore, flow rate conditions exceeding 3.5 m^3/s at St.39 or 2 m^3/s at St. 37 (or St. 38) could have been directly affected by rainfall events.

Figure 4.3 showed the rainfall-runoff observed on different time scale. From the daily river water level (or depth), it could be noticed that the rainfall-runoff response in the study area was quite rapid as seen from short duration narrow peaks. Similarly, from the data of temporal surveys, different runoff patterns were observed depending on the frequency of observation. In case of two days interval observation, majority of periods resembled baseflow even though significant amount of rainfall were observed. In the 10

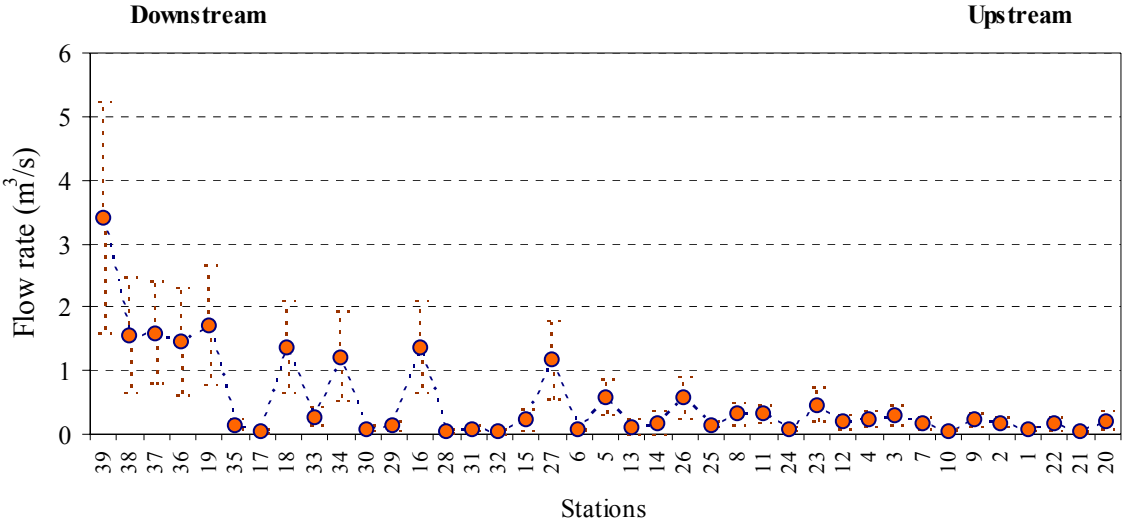


Figure 4.2 Average flow rate at 39 sampling stations

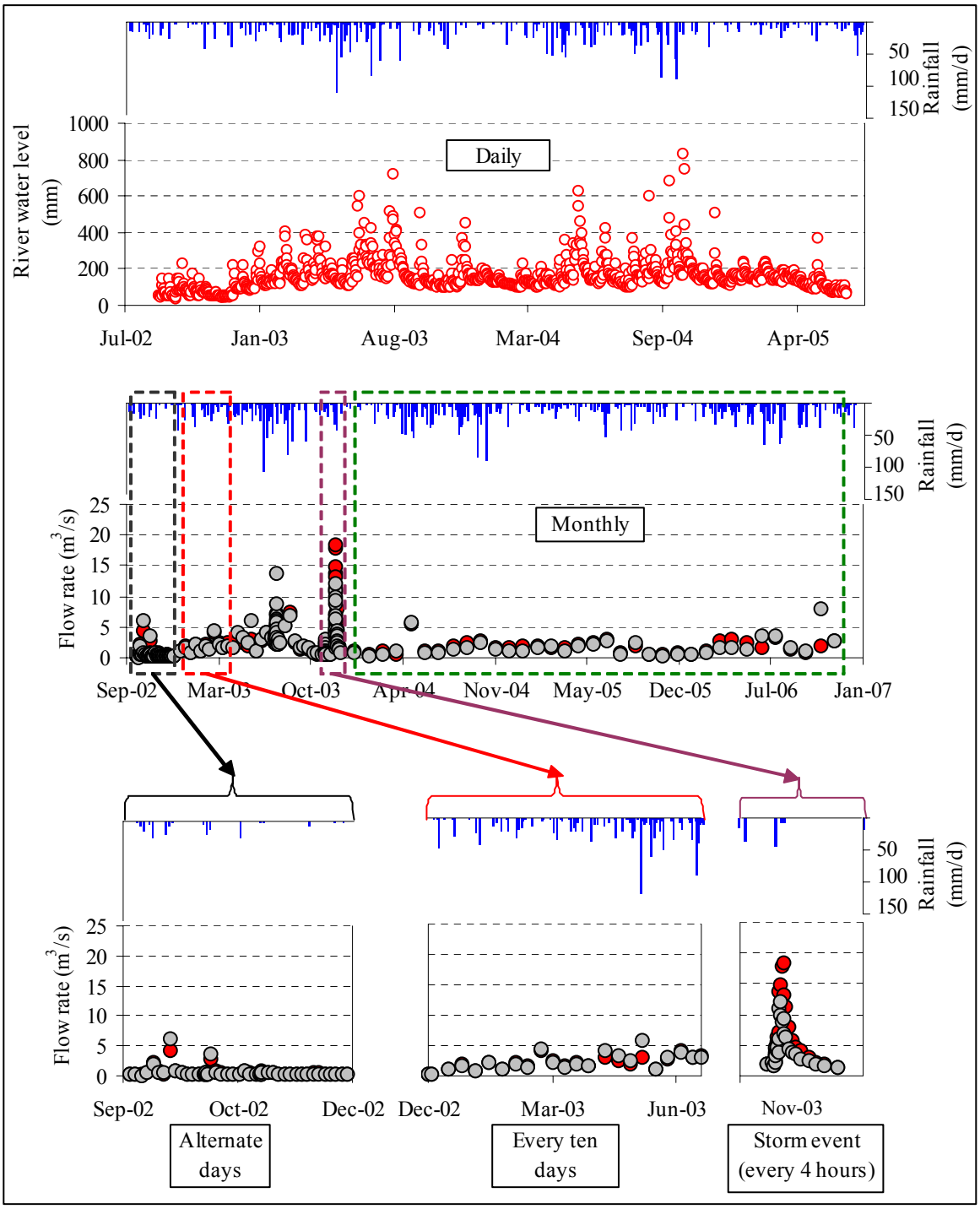


Figure 4.3 River runoff pattern observed at different temporal conditions

days survey, the flow rate was higher than alternate day's survey due to the relatively frequent rainfall events in those days. Because of the rapid rainfall-runoff response, pattern of peak flow event could not be ensured from both types of surveys. The reason could be understood from a storm event survey, in which the highest peak flow rate of 18 m^3/s was observed with cumulative rainfall of 44 mm, but such peaks were never observed in any other types of temporal survey with similar amount of rainfall. Therefore, river runoff could be more dependent on the intensity (or amount) of rainfall at a particular interval of time and they may require frequent observation (~ hourly scale) to understand the peak runoff pattern.

4.3.2 Overview of river water quality

At first, concentration differences of observed WQIs at 39 stations were examined. For example, **Figure 4.4** showed the spatial pattern of average COD_{Mn} and TN concentrations at 39 sampling stations. The average concentration of both indices in the upstream was lower as compared with the downstream. However, some stations near the urban areas showed slightly higher concentrations (St. 17, 25 ~ 33). Average concentrations of other WQIs at different stations are shown in **Appendix B-(1)**, in which the plots are arranged in terms of upstream to downstream. WQIs such as pH, DO, NO_2^--N, SiO_2-Si did not showed detectable variation at stations, but for other WQIs, increasing pattern of the concentrations from upstream to downstream could be clearly observed. WQIs in some stations near the urban areas showed slightly elevated concentrations. It could mean that, urban areas could have some direct influence on the river water quality, most probably from the agricultural fields, from non-sewer areas, or road side runoff (Zhang and Yamada, 1996).

Figure 4.5 showed the concentration of COD_{Mn} and TN observed at the temporal surveys. COD_{Mn} concentration pattern seems to be similar with the flow rate pattern, but TN concentration did not showed, with some exceptions, any big variation with the changes in the runoff conditions. Temporal variation in river water quality was usually affected by the runoff conditions, so it was important to compare the behavior of WQIs at different flow rates. Scattering diagrams showing the relationship between flow rates and observed WQIs are shown in **Appendix B-(2)**. Among the WQIs, three patterns could be observed, namely, WQIs that did not showed much variations (such as pH, DO), WQIs that showed

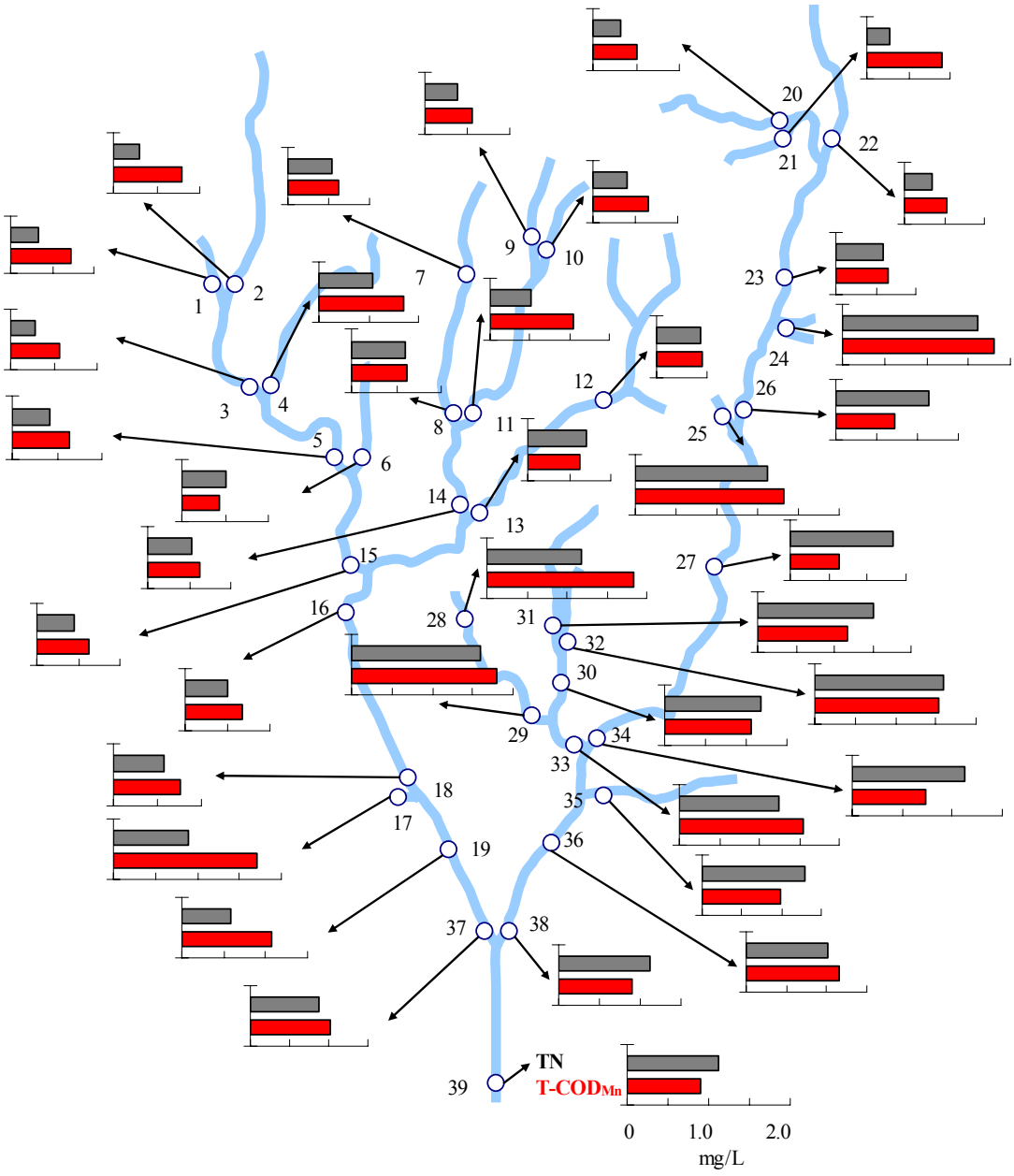


Figure 4.4 Average concentrations of COD_{Mn} and TN at 39 stations

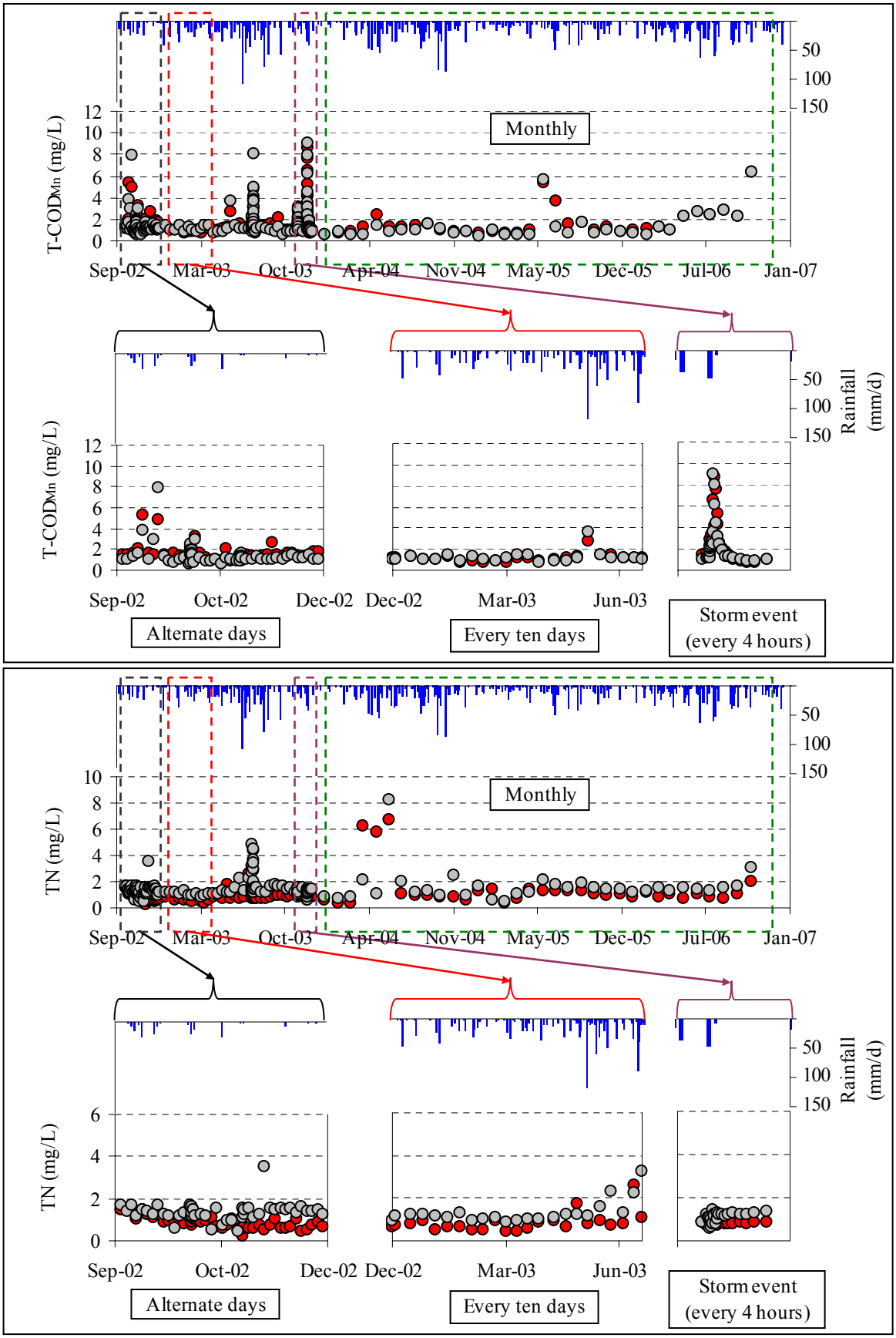


Figure 4.5 COD_{Mn} and TN concentration observed at different temporal surveys

variation but patterns were not clear (such as TN, NO_2^--N, SiO_2-Si, Fe), and WQIs that showed either increasing (such as particulate WQIs) or decreasing pattern (most of the ionic WQIs) with increasing flow rates.

The initial assessment indicated that WQIs could show similar variation pattern at both spatial and temporal conditions. Therefore, further analyses were required, especially, by the use of multivariate analysis techniques that could deal with the multidimensional WQIs.

4.3.3 General statistics of WQIs

Table 4.3 shows the descriptive statistics of all WQIs for both spatial and temporal datasets. Majority of WQIs were below allowable maximum concentration for drinking purpose and aquatic life (Chapman and Kimstach, 1996). It signifies the river was at unpolluted state, which was also indicated by high DO concentration (> 8 mg/L) and nearly neutral pH (6.8 - 8). If we observe the mean concentration of WQIs, some remarks could be made on the values. IC had higher concentration than DOC indicating leaching of carbonate or bicarbonates ions from the watershed. Nearly neutral pH could mean that bicarbonate ions must be higher. Ca^{2+} and Na^+ had the highest concentration among observed cations, whereas SO_4^{2-}, Cl^- and IC were the major anions with higher concentration. Similarly, NO_3^--N had the highest proportion in the total nitrogen.

Concentration range of the major WQIs only hinted the natural state of river water quality and pointed out some key differences, but interrelationship among the WQIs could not be understood from that. Therefore, correlation among the WQIs at both spatial and temporal database was examined and the correlation matrixes of both datasets were provided in the **Appendix B-(3)** and Appendix **B-(4)** respectively. Most of the WQIs were well-correlated ($|r| > 0.5$) among each other in both datasets, including the major anions, cations, nutrients, and organic indices. However, there were some key noticeable similarities or uniqueness of some WQIs in both spatial and temporal datasets. SiO_2-Si was not well-correlated ($|r| < 0.5$) with any WQIs in both datasets. Similarly, majority of WQIs were not well-correlated ($|r| < 0.5$) with: 1) Specific flow rate (SFR), SS, VSS, and Al in the spatial dataset; and 2) nitrogenous indices and Mn^{2+} in the temporal dataset. Inter-comparison of correlation matrixes indicated that WQIs could show significant variation pattern. Therefore, the meaning of correlation pattern was further analyzed to understand

water quality variation occurring in the river.

Table 4.3 Mean values of WQIs of spatial and temporal datasets

WQIs	Spatial dataset (n =234)	Temporal dataset (n =372)
	Mean ± σ (mg/L)	Mean ± σ (mg/L)
pH	7.6 ± 0.5	7.49 ± 0.39
DO	10.5 ± 1.3	9.99 ± 1.68
T-COD_{Mn}	0.92 ± 0.48	1.72 ± 1.31
D-COD_{Mn}	0.68 ± 0.4	1.08 ± 0.57
DOC	0.73 ± 0.35	0.96 ± 0.33
IC	6.93 ± 3.47	7.34 ± 1.79
SS	1.59 ± 1.3	6.48 ± 12.53
VSS	0.64 ± 0.45	1.65 ± 2.85
TN	0.84 ± 0.55	1.32 ± 2.02
DN	0.74 ± 0.49	0.95 ± 0.34
NO_2^--N	0.006 ± 0.007	0.004 ± 0.004
NO_3^--N	0.53 ± 0.36	0.90 ± 0.34
NH_4^+-N	0.02 ± 0.022	0.032 ± 0.076
TP	0.031 ± 0.021	0.047 ± 0.041
DP	0.025 ± 0.017	0.032 ± 0.018
Cl^-	5.43 ± 2.78	7.66 3.04
SO_4^{2-}	7.56 ± 3.42	9.47 ± 1.94
Na^+	5.16 ± 2.4	6.43 ± 2.56
K^+	0.98 ± 0.62	1.14 ± 0.42
Mg^{2+}	1.78 ± 0.63	1.91 ± 0.62
Ca^{2+}	9.34 ± 5.13	10.88 ± 2.97
Sr^{2+}	0.06 ± 0.02	0.07 ± 0.03
Ba^{2+}	0.014 ± 0.007	0.015 ± 0.006
SiO_2	4.64 ± 1.06	3.76 ± 0.80
Fe	0.07 ± 0.08	0.04 ± 0.02
Al	0.007 ± 0.005	0.013 ± 0.014
Mn^{2+}	0.003 ± 0.005	0.002 ± 0.002
SFR#	2.33 ± 1.66	2.81 ± 3.26

Specific flow rate (mm/d)

4.3.4 Spatial dataset

Spatial data were collected from 39 sampling stations at different dates. It was necessary to examine the influence of both sampling dates and stations on the observed variation of WQIs. As an initial assessment, the variation caused by sampling dates and sampling stations on WQIs was examined with two-ways layout ANOVA. The results were shown in the **Fig. 4.6**. Majority of WQIs showed strong effect of stations which was

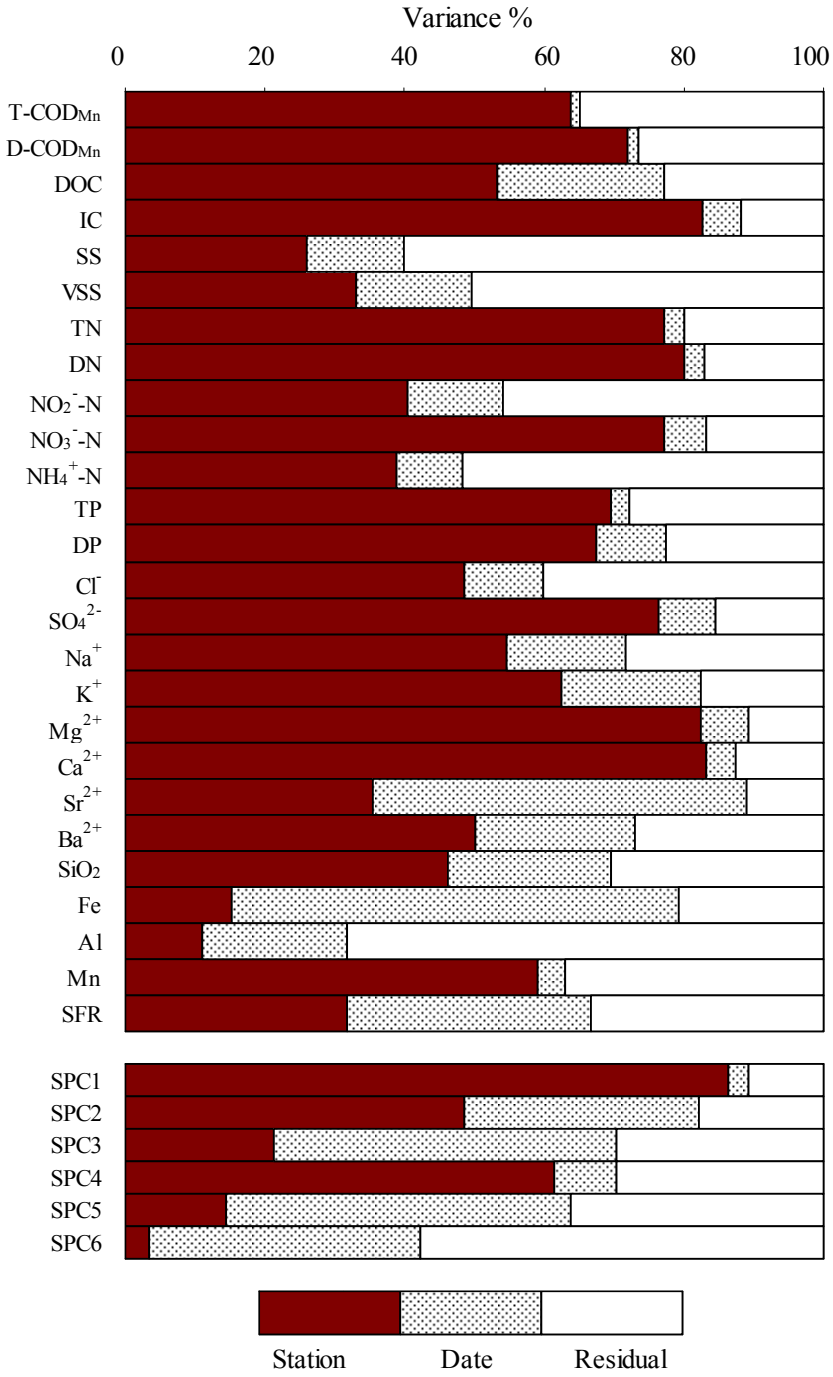


Figure 4.6 Two-ways layout ANOVA for the station and date variation of WQIs and SPCs

higher than that of dates on their variation (≥ 50 %), but WQIs, such as SS, VSS, NO_2^--N, NH_4^+-N, Sr^{2+}, Fe, Al, and SFR, showed a weak effect of stations.

As ANOVA showed that majority of WQIs showed strong effects of stations on their variation, so it was essential to classify stations showing similar water quality. Hierarchical

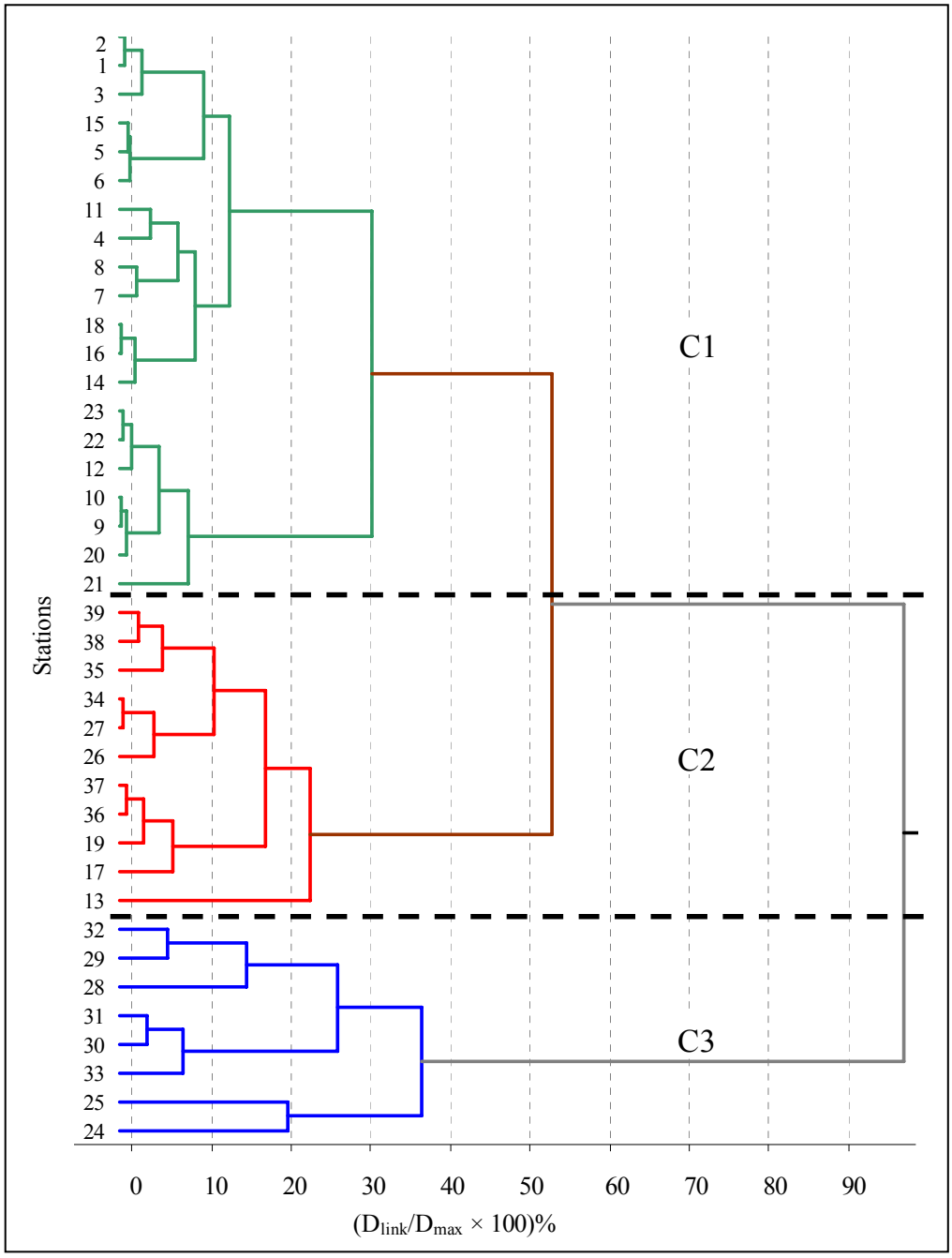


Figure 4.7 Identified clusters with similar water quality variation among 39 sampling stations

agglomerative clustering (CA) was employed to classify the stations into identifiable groups. For this study, we used squared Euclidean distance as a measure of similarity, where as the Wards method was used as clustering criteria. **Figure 4.7** shows the dendrogram for sampling stations, in which D_{link} was a linkage distance for a cluster and D_{max} was the distance of the last clustering. Although variation among identified clusters

Table 4.4 PC loadings (= correlation) for each WQI and explained variance (spatial dataset)

WQIs	SPC1	SPC2	SPC3	SPC4	SPC5	SPC6
T-COD_{Mn}	**0.75**	0.41	-0.12	-0.21	-0.19	-0.03
D-COD_{Mn}	**0.81**	0.26	-0.15	-0.2	-0.15	-0.02
DOC	**0.69**	0.25	-0.34	-0.26	0.09	-0.29
IC	**0.81**	-0.38	0.07	-0.24	-0.1	0.07
SS	0.39	**0.65**	-0.34	-0.13	-0.2	0.17
VSS	0.36	**0.72**	-0.2	-0.06	-0.21	0.07
TN	**0.79**	0.24	0.1	0.35	0.31	0.11
DN	**0.79**	0.25	0.11	0.37	0.28	0.11
NO_2^--N	**0.65**	0.1	0.36	-0.03	-0.25	-0.38
NO_3^--N	**0.77**	0.09	0.16	0.49	0.13	0.05
NH_4^+-N	0.45	0.45	0.21	-0.3	0.29	-0.3
TP	**0.78**	0.18	0.31	0.21	-0.03	0.06
DP	**0.76**	0.06	0.34	0.26	-0.1	-0.01
Cl^-	**0.68**	0.01	-0.43	0.36	0.03	-0.07
SO_4^{2-}	**0.77**	-0.36	-0.15	-0.07	-0.06	0.18
Na^+	**0.82**	-0.21	-0.25	0.15	-0.01	-0.05
K^+	**0.85**	-0.1	-0.12	0.08	-0.2	-0.2
Mg^{2+}	**0.75**	-0.4	0.003	0.07	0.04	0.13
Ca^{2+}	**0.84**	-0.36	-0.05	-0.21	0.04	0.05
Sr^{2+}	0.54	**-0.61**	-0.23	-0.04	0.15	0.08
Ba^{2+}	**0.6**	-0.2	0.16	-0.28	0.3	-0.07
SiO_2	0.27	0.04	**0.6**	-0.48	0.3	0.18
Fe	0.36	-0.21	0.46	-0.03	**-0.6**	0.38
Al	0.18	0.34	-0.32	-0.31	0.22	**0.6**
Mn^{2+}	**0.7**	-0.11	0.14	-0.34	-0.03	-0.17
SFR	-0.26	**0.6**	0.42	0.12	0.02	0.08
Eigenvalue	11.5	3.2	1.9	1.7	1.2	1.1
Variance (%)	44.3	12.3	7.6	6.4	4.5	4.1
Cumulative (%)	44.3	56.6	64.2	70.6	75.1	79.2

Bold face loadings were considered significant ($|r| > 0.6$)

($D_{link}/D_{max}\times 100$) was not so high (< 45 %), sampling stations could be grouped into three clusters (C1 ~ C3). C1 was comprised of stations (St.1~12, 14~16, 18, 20~23) at the upstream, mainly receiving influences from the forest areas. C2 represented the stations that were either near the outlet (St.17, 19, 34~39) or at upstream that have some residential areas, mainly without sewer service, and some paddy fields (St.13, 26, 27), while C3 was comprised of the stations receiving major urban influences (St.24, 25, 28~33).

Both ANOVA and CA indicated that majority of WQIs could show spatial variation pattern. PCA was then applied to the spatial dataset to synthesize the variation patterns of WQIs at different stations and its summary is shown in **Table 4.4**. Six principle components (PCs), *i.e.*, "spatial principle components (SPCs)", were identified as

significant (eigenvalue > 1), which explained nearly 79% of variance of the dataset. However, only SPC1 (44% of variance) and SPC2 (12%) seemed to be important in terms of WQIs represented by them. SPC1 had significant loadings ($|r| > 0.6$) for majority (~69%) of WQIs, while SPC2 showed significant **loadings** only with SFR, SS, VSS, and Sr^{2+}. Remaining SPCs contributed less than 10% of variance individually and represented only a few WQIs.

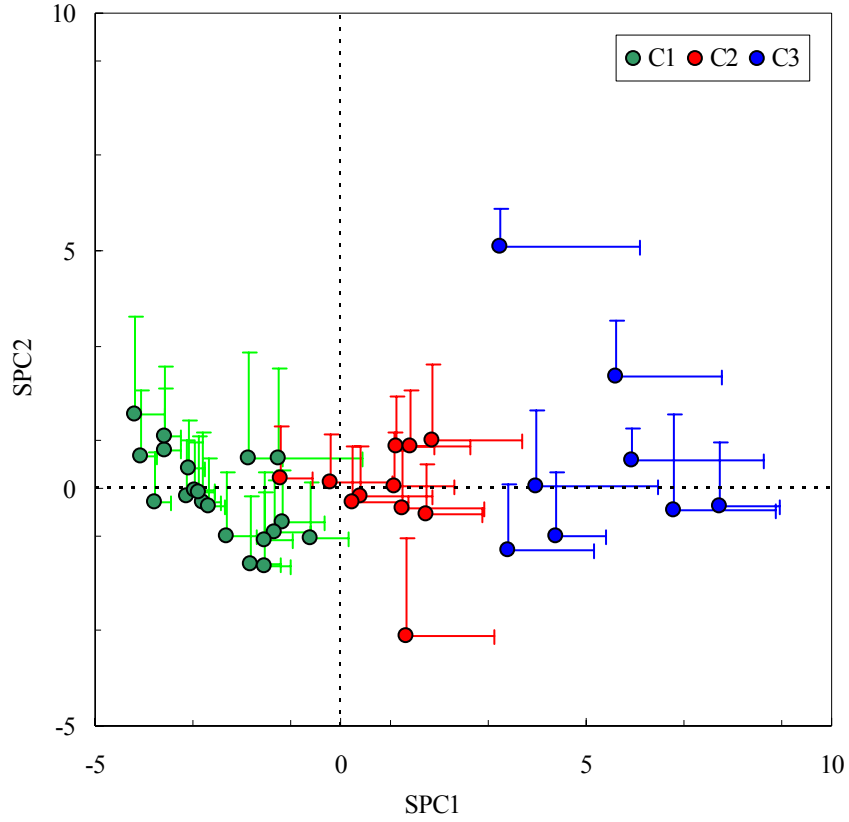


Figure 4.8 SPC1 and SPC2 mean scores and standard deviation (+σ) at sampling stations with respect to identified clusters

Since SPCs representing spatial patterns were important in this analysis, two-ways layout ANOVA, for sampling stations and dates variations, was also applied to the scores of significant SPCs and results are shown in the Fig. 4.6. SPC1 showed the highest variation among the WQIs due to stations, whereas SPC2 also showed variations due to dates. Similarly, SPC1 and SPC2 scores were also compared with respect to identified clusters as shown in **Fig. 4.8**. Dots in Fig. 4.8 are averages of SPC scores for each station and bars are standard deviations (+σ). The patterns of SPC2 scores could not be easily distinguished but SPC1 distinctly separated three clusters. C1 and C2 were closer and less deviated to SPC1 axis, while C3 was more scattered and deviated. Since SPC1 was

positively correlated with majority of WQIs, stations in C1 and C2 mainly represented lower concentration while stations in C3 represented relatively higher concentration of those WQIs. SPC1 showed the spatial variation pattern and represented majority of WQIs, so it will be discussed further in latter section.

4.3.5 Temporal dataset

As mentioned earlier, the rainfall runoff response in the study area was very rapid with the duration of peak runoff less than two days in most of the cases. Assuming runoff had major impact on the temporal variation of river water quality, the temporal dataset was classified according to the quartile distribution (Q1 ~ Q4) of SFR (mm/d). The quartile ranges of SFR were as follows: Q1: 0.1 ~ 0.6; Q2: 0.6 ~ 1.74; Q3: 1.74 ~ 3.65; Q4: 3.65 ~ 20.96.

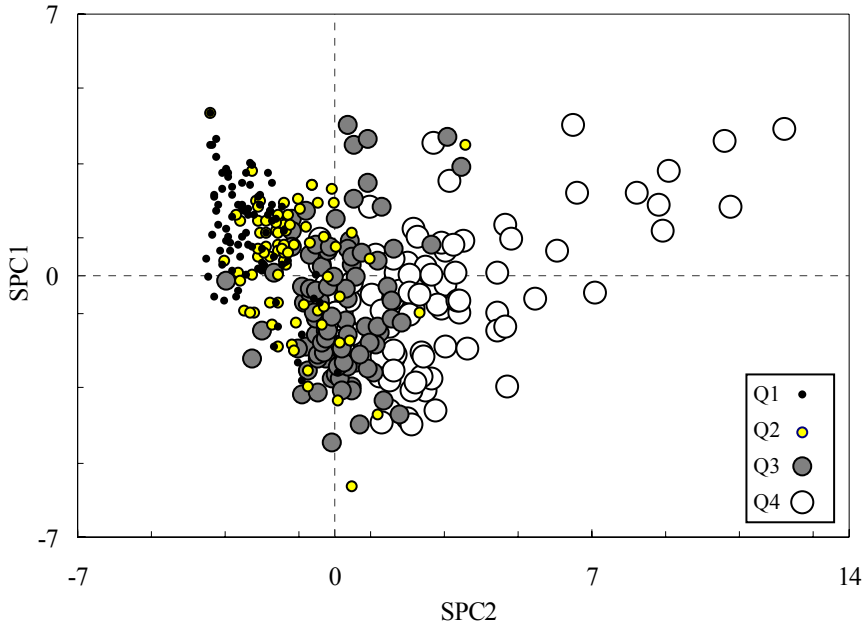


Figure 4.9 SPC1 and SPC2 scores of the temporal dataset derived from eigenvectors of spatial PCA

Since observed WQIs in both spatial and temporal datasets were same, applicability of eigenvectors ($a_{i1}...a_{ij}...a_{im}$) (**Eq. 2.3**), obtained in "spatial PCA", were examined in the temporal dataset. **Figure 4.9** shows the scattering diagrams for SPC1 and SPC2 scores of temporal dataset that was derived from the eigenvectors obtained in "spatial PCA". As majority of WQIs represented by the SPC1 were only positively correlated, the scores pattern could not be easily interpreted. However, patterns of SPC2 scores could be easily

identified. Higher magnitude scores (towards positive SPC2) were mainly represented by Q3 ~ Q4, while the lower magnitude scores (towards negative) were mainly represented by the Q1 ~ Q2. The patterns of SPC2 scores might indicate the sensitivity of significant WQIs (e.g., SS, VSS, Sr^{2+}, SFR in Table 4.4) to the rainfall-runoff conditions.

Due to the less variation of SFR in the spatial dataset, PCA was applied to the temporal dataset to generalize the water quality variation patterns at different runoff conditions. PCA explained more than 75% of total variance of the temporal dataset by five TPCs (temporal PCs) (**Table 4.5**). Among TPCs, only first three TPCs seemed to be more informative in terms of significant loadings WQIs. TPC1 (35.8 % of variance) had majority WQIs with significant loadings, but TPC2 (17.2 %) had only K^+, but majority of other WQIs showed positive moderate loadings ($0.6 > r > 0.5$). TPC3 (8.3%) was represented mainly by DN and NO_3^--N. Variation patterns of TPC1, TPC2, and TPC3 scores were then examined by using scattering diagrams as shown in **Fig. 4.10**.

Although PCs are uncorrelated, TPC1 and TPC2 patterns seemed to show inverse and direct relationship with each other for the group of scores representing the SFR ranges of "Q1~Q2" and "Q3~Q4" respectively. Increasing SFR (Q1 to Q4) also increased the level of TPC1 scores, whereas increasing SFR firstly decreased the level of scores (Q2) of TPC2, but later on (> Q2) scores started to increase proportionally. In case of TPC3 (TPC1 and TPC3), scores variation were seen decreasing with increasing SFR, but for higher quartile range (> Q3) scores showed nearly constant level.

4.3.6 Effect of land use on river water quality

Majority of WQIs showed distinct variation pattern at different stations. PCA on the spatial dataset showed that SPC1 was correlated ($|r| > 0.6$) with majority of WQIs and represented the spatial variation pattern (Table 4.4). It could mean that majority of these WQIs generally covary together at different stations. However, it was necessary to understand whether the covariance of majority of WQIs was due to the influence of natural or anthropogenic factors. The patterns of identified clusters could mean that river water quality at the upstream was influenced by the forest areas (stations in C1) but near urban and agricultural areas river water quality was also affected by the human activities (stations in C3). However, near the outlet (majority of stations in C2), the concentration of majority WQIs were influenced by both forest and urban areas. Since majority of population (>

90%) in the densely populated urban areas were well connected with sewer service, so their loadings were transmitted outside of the watershed. Similarly, if we consider lesser proportion of urban (~ 11% of area) and agricultural (3%) areas (with respect to digitized land use map), and less variation among identified clusters ($D_{link}/D_{max}\times100 < 45$ %), impacts of human activities on the river water quality could be lesser as compared with the dominant forest areas.

Table 4.5 PC loadings (= correlation) for each WQI and explained variance (temporal dataset)

WQIs	TPC1	TPC2	TPC3	TPC4	TPC5
T-COD_{Mn}	**0.73**	0.57	-0.03	-0.23	0.10
D-COD_{Mn}	0.57	0.35	-0.08	-0.21	0.12
DOC	**0.67**	0.41	0.00	0.07	-0.12
IC	**-0.69**	0.30	-0.14	-0.07	0.01
SS	**0.72**	0.54	-0.03	-0.25	0.03
VSS	**0.68**	0.55	-0.04	-0.20	-0.04
TN	0.12	-0.06	0.54	-0.03	-0.36
DN	-0.11	0.17	**0.85**	0.20	-0.14
NO_2^--N	0.17	0.29	-0.03	0.29	-0.71
NO_3^--N	-0.39	0.32	**0.69**	-0.06	0.26
NH_4^+-N	0.59	0.10	-0.01	0.00	-0.40
TP	**0.76**	0.56	0.13	-0.03	0.03
DP	**0.62**	0.41	0.36	0.18	0.05
Cl^-	**-0.64**	0.29	0.21	-0.21	-0.20
SO_4^{2-}	**-0.77**	0.26	0.28	-0.06	-0.02
Na^+	**-0.76**	0.51	0.00	-0.10	0.03
K^+	-0.52	**0.75**	0.14	-0.02	0.12
Mg^{2+}	**-0.81**	0.47	-0.02	0.06	0.08
Ca^{2+}	**-0.76**	0.54	-0.16	0.08	0.08
Sr^{2+}	**-0.71**	0.39	-0.20	0.02	-0.12
Ba^{2+}	-0.37	0.54	-0.25	0.24	-0.08
SiO_2	0.14	-0.41	0.45	0.37	0.38
Fe	0.05	0.46	-0.29	0.48	0.26
Al	**0.69**	0.32	-0.09	0.23	0.19
Mn^{2+}	0.28	0.13	-0.09	**0.75**	-0.03
SFR	**0.86**	0.15	0.10	-0.14	0.18
Eigenvalue	9.3	4.5	2.2	1.5	1.3
Variance (%)	35.8	17.2	8.6	5.8	5.0
Cumulative (%)	35.8	53.0	61.6	67.4	72.4

Bold face loadings were considered significant ($|r| > 0.6$)

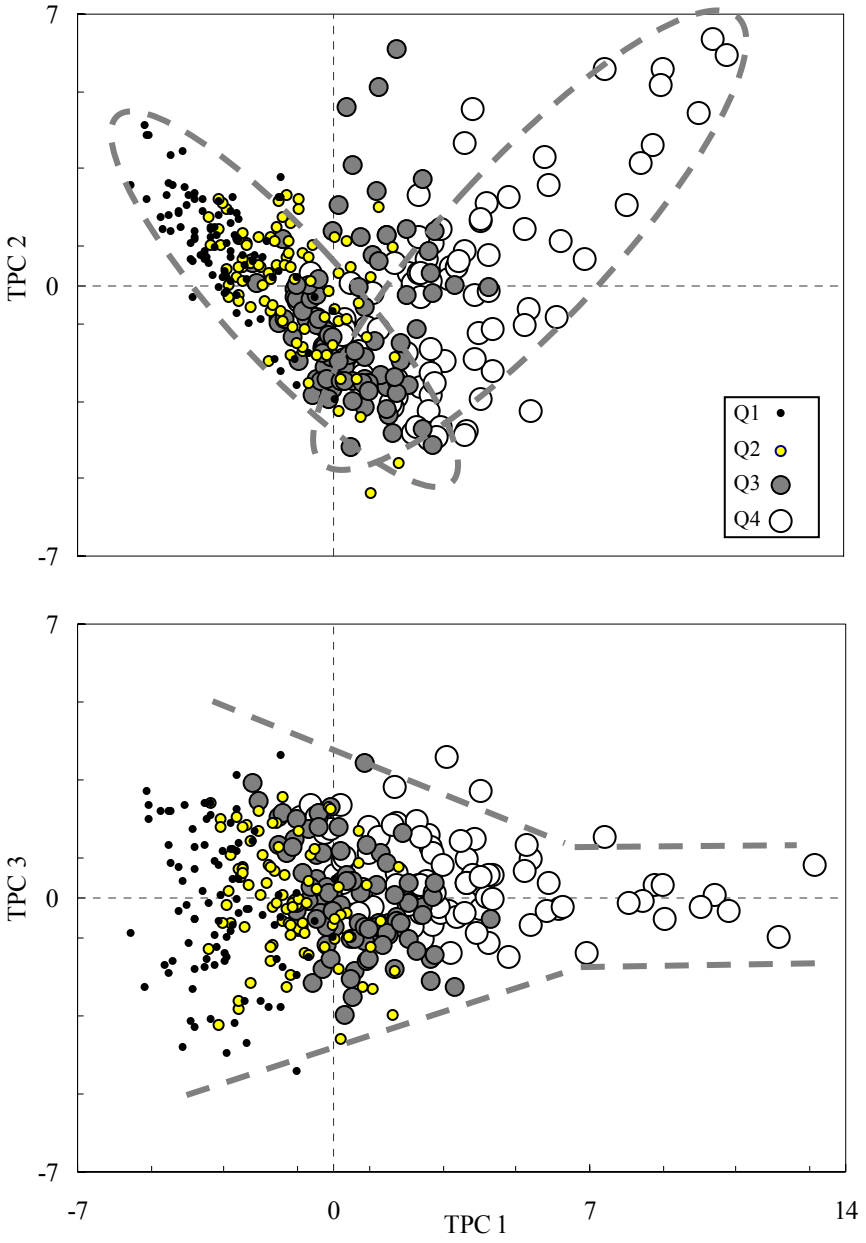


Figure 4.10 TPCs scores patterns at different SFR ranges obtained after applying PCA to the temporal dataset

4.3.7 Effect of runoff conditions on river water quality

Rainfall-runoff conditions could have significant influence on the river water quality. Therefore, it was quite important to understand the covariance pattern of WQIs at different runoff conditions. WQIs showing similar covariance patterns in temporal dataset could be grouped based on significant positive or negative loadings (*i.e.*, correlation) with TPCs. WQIs in TPC1 (Table 4.5) could be roughly divided into two groups: 1) significant

negative loadings ($r < -0.6$) (*Type I*: IC, Cl^-, SO_4^{2-}, Na^+, K^+ (-0.52), Mg^{2+}, Ca^{2+}, Sr^{2+}); and 2) significant positive loadings ($r > 0.6$) (*Type II*: T-COD_{Mn}, DOC, SS, VSS, TP, DP, NH_4^+-N(0.59), Al, and SFR) WQIs. Similarly, it could also be noticed that TPC2 showed only positive correlation with most WQIs [*e.g.*, T-COD_{Mn} (0.57), SS or VSS (0.55), TP (0.56), Na^+ (0.51), K^+ (0.75), Ca^{2+} (0.54), and Ba^{2+} (0.54)]. Therefore in Fig. 4.10 (TPC1 versus TPC2), scores of "Q1" represented the condition with higher concentration of "*Type I*" WQIs, which mostly constitutes the dissolved WQIs, but gradual increase in the SFR (Q2 to Q4) decreased their concentration. However, increasing SFR also increased the concentration of "*Type II*" WQIs, which mostly constitutes the particulate related WQIs. It could be assumed that "*Type I*" WQIs had limited rate of release, so their concentration declined with increasing runoff volume (from both surface and underground) due to the dilution effect. In case of "*Type II*" WQIs, they could be available abundantly whenever surface runoff occurred. It could be assumed that surface runoff were possible only during heavy rainfall events so that natural accumulation of "*Type II*" WQIs in the watershed were not affected by smaller rainfall.

TPC3 was significantly positively correlated (loadings > 0.6) with only DN and NO_3^--N. Again in Fig. 4.10 (TPC1 versus TPC3), variation of the TPC3 scores during lower SFR conditions (Q1 ~ Q2) could be due to the differences in the seasonal nitrogen demand by the plants, climatic conditions that might affect the rate of mineralization of nitrogenous compounds, or due to the antecedent conditions. However, during higher SFR conditions, either input from the rainwater or higher availability of nitrogen in the watershed could cause constant NO_3^--N concentration in the river. Ohuri and Mitchell (1997) had mentioned that when nitrate level in the river did not show any seasonal variation in forested watersheds of Japan, it could be an indicator of nitrogen saturation condition. Nitrogen saturation conditions were generally indicated by the bulk precipitation of nitrogen greater than 10 kg N ha^{-1} yr^{-1} (Wright et al., 1995), which is quite close to the range of mean bulk deposition of N (3.5 - 10.5 kg N ha^{-1} yr^{-1}) observed at different forested watersheds inside Japan (Ohrui and Mitchell, 1997; Mitchell *et al*., 1997; Ohte *et al*., 2001b). However, in our study, mean concentration of NO_3^--N observed in the river water at both spatial and temporal datasets were quite lower, but it could be possible that nitrogen in the soils of forest areas were very high.

4.3.8 Role of forest areas on the spatial and temporal variations of WQIs

Foregoing analysis of the spatial dataset indicated that role of forest areas on the river water quality could be dominant compared to the limited anthropogenic influences (from urban areas and agricultural practices). Similarly, scores pattern of TPC3 at higher SFR also indicated the possibility of nitrogen saturation condition in the forest areas. Therefore, it was quite imperative to analyze the role of forest areas on the variation pattern of majority of WQIs in both spatial and temporal contexts. Particularly, possible inter-relationships between the rainfall (quantity and quality) and biogeochemical processes occurring inside the forest ecosystem will be discussed.

Rainfall and resultant runoff at surface, subsurface and underground are important for both biogeochemical reactions and mobility of WQIs. Therefore, impacts of rainfall on the biogeochemical reactions and resultant river water quality should be analyzed at first. In case of spatial surveys, they were scheduled to avoid direct influence of rainfall, but for most of the times sufficient moisture was supposed to be available on the subsurface layer of the watershed due to frequent rainfall in the study area (Fig. 3.6). The fact could be realized from the large number of SFR observations (n = 183) in the temporal dataset that were less than the mean SFR (< 1.88 mm/d) at same stations (St. 37 and St. 38) in the spatial dataset. It could mean that percolation or subsurface runoff could be occurring, possibly at relatively slower rate, for considerable period even after the cessation of peak runoff events. In such conditions river water quality could be affected not only from the underground but also from the subsurface layer. Tsujimura *et al*., (2001) had also found that, in a steep and high relief basin of central Japan, stream water chemistry was mostly influenced by the subsurface water flowing through the bedrock. Besides, distinct variation patterns of "*Type I*" and "*Type II*" WQIs, which mainly constitutes the dissolved and particulate related WQIs respectively, also indicated different role of surface and subsurface (or underground) runoff processes on WQIs variation in our study area. In a forested watershed of central Japan, Ohrui and Mitchell (1999) have found that during smaller rainfall events (or without rainfall) (< 100 mm in 13 ~ 25 hours) mainly groundwater or subsurface flow (lesser proportion) contributed stream water chemistry. Whereas during larger rainfall events (> 100 mm in 14 ~ 24 hours), subsurface flow was the main contributor of stream water chemistry. Due to majority of steep hilly areas and relatively shallower soil profiles (< 3m ~ 10m) in our study area, subsurface hydrology

could be the major contributor of river water quality during most of the times, however during higher rainfall periods surface runoff could also affect the river water quality. However, contribution of surface, subsurface, or underground layer on river water quality should be distinguished in terms of biogeochemical processes occurring inside the forest areas.

It has been already shown that, with few exceptions, majority of dissolved as well as particulate related WQIs generally covary together at both spatial and temporal conditions. Which means biogeochemical processes occurring inside the forest ecosystem could be interlinked. Muraoka and Hirata (1988) had mentioned that biogeochemical reactions inside the soil profile could be the main determinant of element cycle of forest ecosystem, which in turn could be deeply linked with the river water quality. One of the important factors related with the physico-chemical reactions occurring inside the forest ecosystem is the impact of acid rainfall, which is a serious environmental problem in Japan. Average pH of rainwater in Japan was reported in the range of 4.7 - 4.9, where as in Kyoto prefecture, the average pH was reported in the range of 4.5 - 4.7 (MOE, 2003). Acidic rainfall, besides being the source of nitrogen, is also believed as the prime cause of rapid rate of rock weathering and higher base cation exchange in the soils resulting nearly neutral to alkaline (pH: 6.1-8.1) stream water (Nakamura *et al*., 1984; Ikeda and Miyanaga, 1995; Asano and Uchida, 2005). According to Asano and Uchida (2005), major portion of neutralization of acidic rainfall could occur in the soil profile when soil was base saturated, but it could be through rock weathering when the base cation in the overlying soil profile were already leached. Major soils in the forested watersheds of Japan, including our study area are Cambisols, which are usually characterized by low amount of weathered minerals (because of soil erosion and landslides) and higher base saturation (Ohte *et al*., 2001a). Therefore, covariance of major cations (*e.g*., Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}) and anions (NO_3^--N, SO_4, and IC) observed in the spatial dataset may indicate the base cation exchange as major physico-chemical processes in the soil profile of the forest areas. However, in temporal dataset, absence of any correlation ($|r| < 0.5$) between NO_3^--N and major base cations (*e.g*., Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}), could mean that NO_3^--N may not have significant role on the base cation exchange. It could be realized if we compare the mean concentration of Ca^{2+} and major anions (NO_3^--N, SO_4^{2-}, and IC), in which concentration of NO_3^--N was the lowest.

It was also equally important to understand whether the sources of base cations were

from soils or through rock weathering. Geologically, watershed seems to be uniform, mostly composed of sedimentary rocks that usually have higher silica content. Therefore, rock weathering could supply more SiO_2 than other elements. SiO_2 is considered as an indicator of rock weathering and generally unrelated with significant metabolic activities. Its concentration in the river possibly depends on the contact time between groundwater and rocks (Ikeda and Miyanaga, 1995; White and Blum, 1995; Tsujimura *et al*., 2001; Asano *et al*., 2003). In several watersheds of Japan, SiO_2 was found to covary with major ions (Ca^{2+}, Mg^{2+}, K^{+}, SO_4^{2-}) in the river, for which geology is often believed to be the main source(Ohrui and Mitchell, 1999; Anazawa and Ohmori, 2001, Tsujimura *et al*., 2001). However, in our study, SiO_2-Si was not well-correlated with any WQIs, but major ions (Ca^{2+}, Mg^{2+}, K^{+}, Ba^{2+}, Sr^{2+}, SO_4^{2-}) and several other WQIs (IC, DOC, T-COD_{Mn}, D-COD_{Mn}, TP, DP *etc*.) were well-correlated among each other in both spatial and temporal datasets. Therefore, rather than from rock weathering, surface and subsurface processes could have dominant effect on the river water quality. However, in future, further verification of sources of SiO_2 in the study area could be helpful to support the foregoing discussions.

4.3.9 Relevance of identified variation patterns on the river water quantity and quality modeling

Rest part of this dissertation is contributed towards distributed water quantity and quality modeling, so the identified water variation patterns could form the basis to set the hypothesis during the design and parameterization of the distributed model. In river runoff modeling, it was quite important to consider runoff variability cause by the surface and sub-surface flow processes. Especially, influence of steep topography and shallower soil profiles could exert dominant influence on the surface and sub-surface runoff process after the initiation of rainfall events. Besides, it was quite evident that due to limited influences from the human activities and dominant influence of forest areas, major variation on the model performance would be influenced by the parameters of forest areas. However, effects of urban areas could be observed on the rapid accumulation of water due to relatively impermeable surface and lower slope.

Analyses of water quality data also showed three types of WQIs. *Type I* WQIs showed that that their rate of release remains constant irrespective of runoff volume, depicting dilution effect with the increasing runoff. *Type II* WQIs, however, showed

proportionate relationship with the surface runoff. In case of nitrogen, especially inputs from the atmosphere as well as their seasonal changes inside the forest ecosystem should be considered. However, this study was more concerned in demonstrating the potential of distributed model to simulate the variation of river water quality, so nitrogen modeling was not focused in this study. However, depending on the sensitivity on the prediction of each WQIs, complexity of modeling considering specific processes could be accomplished in future. Therefore, the water quantity and quality variations pattern identified in this study is expected to be helpful to conceptualize the river water quality modeling processes primarily focusing on the inter-linkages among the rainfall pattern, hilly topography, and surface and sub-surface hydrology in the forest areas.

4.4 Summary

Important findings related with this chapter could be briefly summarized as:

1) Covariance patterns of majority of WQIs (~ 69%) in both datasets indicated their common sources, of which urban and agricultural areas (< 15 %) could have limited contribution as compared to the dominant forest areas (> 85 %);

2) Rainfall could mainly cause dilution effect on some ionic water quality indices (*e.g.*, IC, Cl^-, SO_4^{2-}, Na^+, K^+, Mg^{2+}, Ca^{2+}, Sr^{2+}, and Ba^{2+}), but it could increase the concentration of particulate water quality indices (*e.g.*, T-COD_{Mn}, D-COD_{Mn}, DOC, SS, VSS, TP, DP, NH_4^+-N, Al) by sweeping effect of resultant surface runoff;

3) Covariance of majority WQIs in both spatial and temporal conditions could be due to biogeochemical processes that seemed to be closely inter-linked among rainfall pattern, deposition of acidic ions, hilly topography, shallower soil profile, base cation exchange in the subsurface, and resulting surface and sub-surface runoff processes in the forest areas;

4) The covariance of majority WQIs and its inter-linkages with the biogeochemical processes in the forest areas indicate the importance of surface and sub-surface runoff on both river water quantity and quality.

CHAPTER V

DEVELOPMENT OF A DISTRIBUTED WATER QUANTITY AND QUALITY MODEL

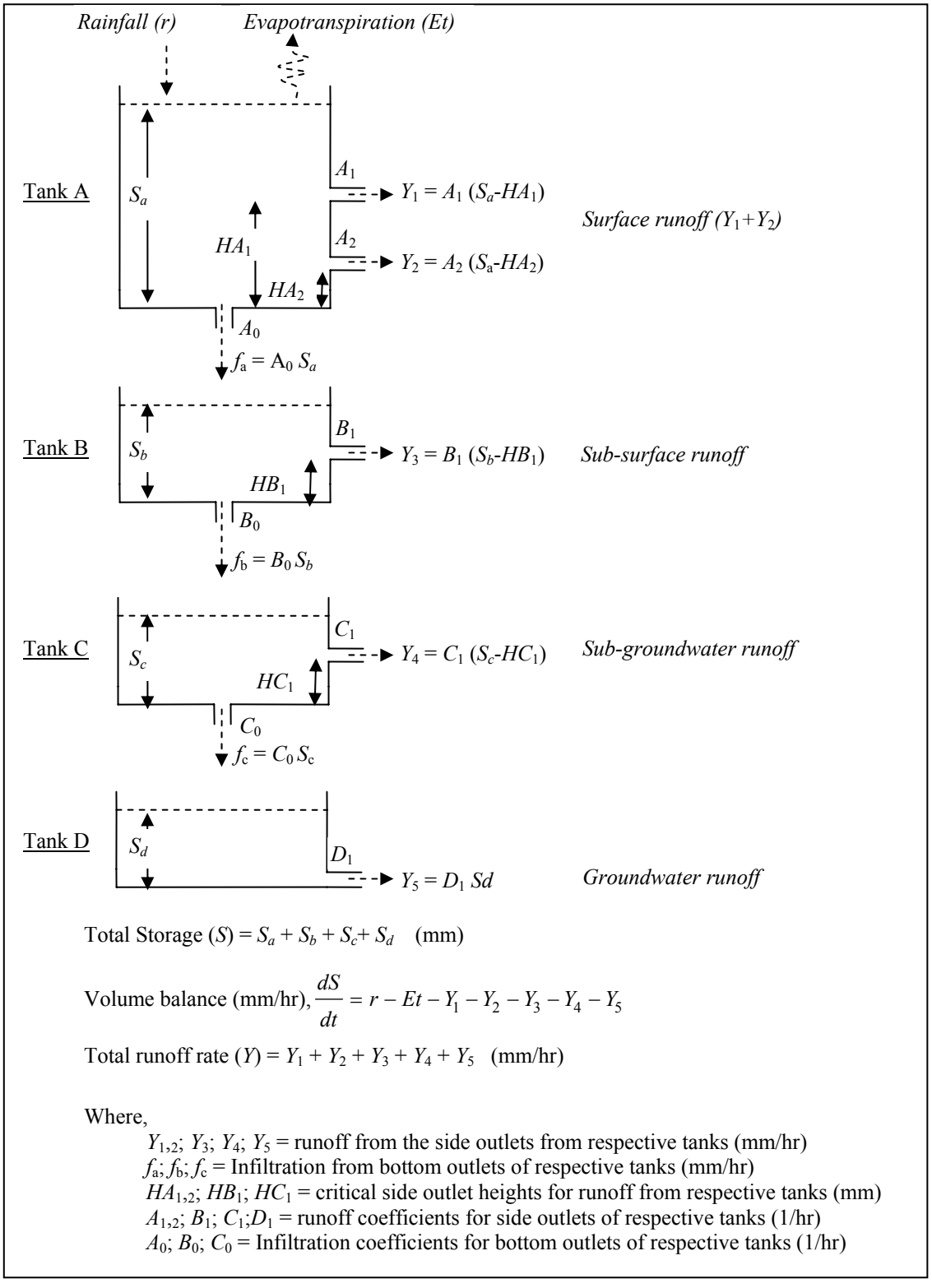
5.1 Introduction

Distributed models are characterized by their capability to conceptualize and incorporate runoff processes occurring at different spatial scale. Different from lumped models, major aim of distributed models is to consider specific differences in the watershed attributes that could possibly affect the hydrological processes. In distributed models, hydrological processes occurring at grid scale are integrated into desired outputs, such as river discharge. Since distributed models need to consider several spatial details, they are quite data intensive. In many instances, most of the data demanded by a distributed model are not available at the desired scale. Such conditions often complicate the modeling processes thereby offsetting the credibility of distributed models.

In this chapter, we constructed a simple distributed model based on the conceptual framework of lumped tank model. Major emphasis was put onto the inputs of the model that could be derived from existing spatial database, while modeling process for each grid was kept as simple as possible.

5.2 Conceptual framework of lumped tank model

Lumped tank model is still popular because of its simple conceptual framework and successful estimation in most of the cases (Sugawara, 1979, 1995; Lee and Singh, 1999, 2005; Hashino *et al*, 2002; Tingsanchali and Gautam, 2000; Setiawan *et al*., 2003). The widely used form of tank model consists of four vertically layered storage tanks in series as shown **Fig. 5.1**. The first tank simulates the surface runoff, second tank simulates the subsurface runoff, and the third and fourth tanks simulate the groundwater runoff. Upper tank also receives the rainfall and lose the water through evapotranspiration. Therefore, storage of the first tank is attributed to rainfall and the storage of the lower tank to the infil-



Total Storage (S) = $S_a + S_b + S_c + S_d$ (mm)

Volume balance (mm/hr), $\frac{dS}{dt} = r - Et - Y_1 - Y_2 - Y_3 - Y_4 - Y_5$

Total runoff rate (Y) = $Y_1 + Y_2 + Y_3 + Y_4 + Y_5$ (mm/hr)

Where,

$Y_{1,2}$; Y_3; Y_4; Y_5 = runoff from the side outlets from respective tanks (mm/hr)
f_a; f_b; f_c = Infiltration from bottom outlets of respective tanks (mm/hr)
$HA_{1,2}$; HB_1; HC_1 = critical side outlet heights for runoff from respective tanks (mm)
$A_{1,2}$; B_1; C_1;D_1 = runoff coefficients for side outlets of respective tanks (1/hr)
A_0; B_0; C_0 = Infiltration coefficients for bottom outlets of respective tanks (1/hr)

Figure 5.1 Conceptual structure of lumped tank model

tration from the upper tank. The runoff rates and infiltration from each outlet is linearly proportional to the storage height above the outlets. If water is not available above any outlets, no output (runoff or infiltration) could occur from those outlets. Although runoff from each tank is linearly proportional, the integral behavior of whole tank model is completely non-linear due to combination of the tank components as well as due to the positioning of side outlets slightly above the bottom of each tank (except for the lowest tank). Such condition will allow modeling of non-linear behavior of runoff processes inside the watershed such as initial loss of the rainfall in the voids or dry soil layer. Similarly, depending on the complexity of hydrological processes, the model could be conceptualized into different forms by changing both the number and position of storage tanks or outlets (Sugawara, 1995).

The main parameters of tank model are the critical heights of side outlets ($HA_{1,2}$; HB_1; HC_1), runoff coefficients($A_{1,2}$; B_1; C_1), and infiltration coefficients (A_0; B_0; C_0) as shown in the Fig 5.1. Appropriate values of these coefficients are generally decided during model calibration or validation, which could be simple to quite complex depending on the hydrological characteristics of the watershed. The height of the outlets reflects the physical concept of the runoff process, such as infiltration of the first tank start immediately with rainfall but surface runoff is delayed until the water storage in the upper tank is above the side outlet (Lee and Singh, 2005). For the lower tanks, similar runoff processes occur from the outlets. Whereas the runoff coefficients represents the different factors (such as slope, surface roughness, or conductivity of underground) affecting the runoff. The total runoff is computed by adding the runoff from each side outlet.

5.3 Transformation of tank model (lumped → distributed)

Lumped tank model is quite simple in structure, which conceptualizes the physical meaning of three major hydrological processes (surface, sub-surface, and groundwater) in terms of the storage balance in a cascade of tanks. In spite of its simple conceptual framework, tank model has been found successful to reflect the phenomenon of surface runoff, sub-surface runoff, and baseflow in variety of runoff conditions ranging from flood to low flows. However, lacking consideration of hydrological processes due to the spatial variability inside the watershed and lacking of specific criteria, other than calibration, to assign the values of major parameters are the major limitations of lumped tank model.

Different from the lumped models, distributed models aim to incorporate all hydrological processes due to the spatial variability in terms of governing equations so that the models are more representative to the major factors affecting the hydrological processes. In distributed models, the whole watershed area is divided into small units or grids (usually rectangular shape) and hydrological processes inside each grid are independently modeled so that models are more flexible in designing and application. In distributed model, each grid could be based on the mechanistic approach, in which the equations of conservation of mass and energy are explicitly applied for each grid, such as MIKE-SHE model. As another recent approach, hydrological responses could be based on the variable contributing area concept, which is the fundamental concept behind TOPMODEL. Another type of distributed modeling approach could be thought as storage type distributed model, such as distributed tank model (Yoshino *et al.*, 1991).

Spatial information of each grid is one of the major requirements of distributed models and performance of the model largely depends on the availability of spatial details (or distributed information). In reality, except few cases, all of the information in each grid required by a distributed is rarely available in the specified scale, even though the model is physically sound. Then, one has to rely on or extrapolate the available data that are available only at certain points (such as rainfall, soil type, or geology).

Simpler types of distributed models requiring limited amount of spatial details are therefore more desirable than the complex models that demand a lot of unknown distributed information to assign the values of its parameters. One of such simple distributed model could be the distributed tank model, which is conceptually similar to the lumped tank model. In case of distributed tank model, instead of whole watershed, vertically layered tanks are conceptualized for each grid and values of parameters are decided based on the spatial variability rather than only calibration tasks. The grids are interconnected on the basis of routing technique that is used to transfer the outflows from the upstream to the downstream grid at a given time. For distributed tank model, first major requirement is to utilize the spatial information that is readily available (such as land cover, geology, DEM). Second requirement is to reduce or simplify the number and positioning of tanks or outlet heights in each grid in order to minimize the number of unknown parameters. It would be more appropriate if the number of tanks or outlet heights were decided on the basis of available distributed information.

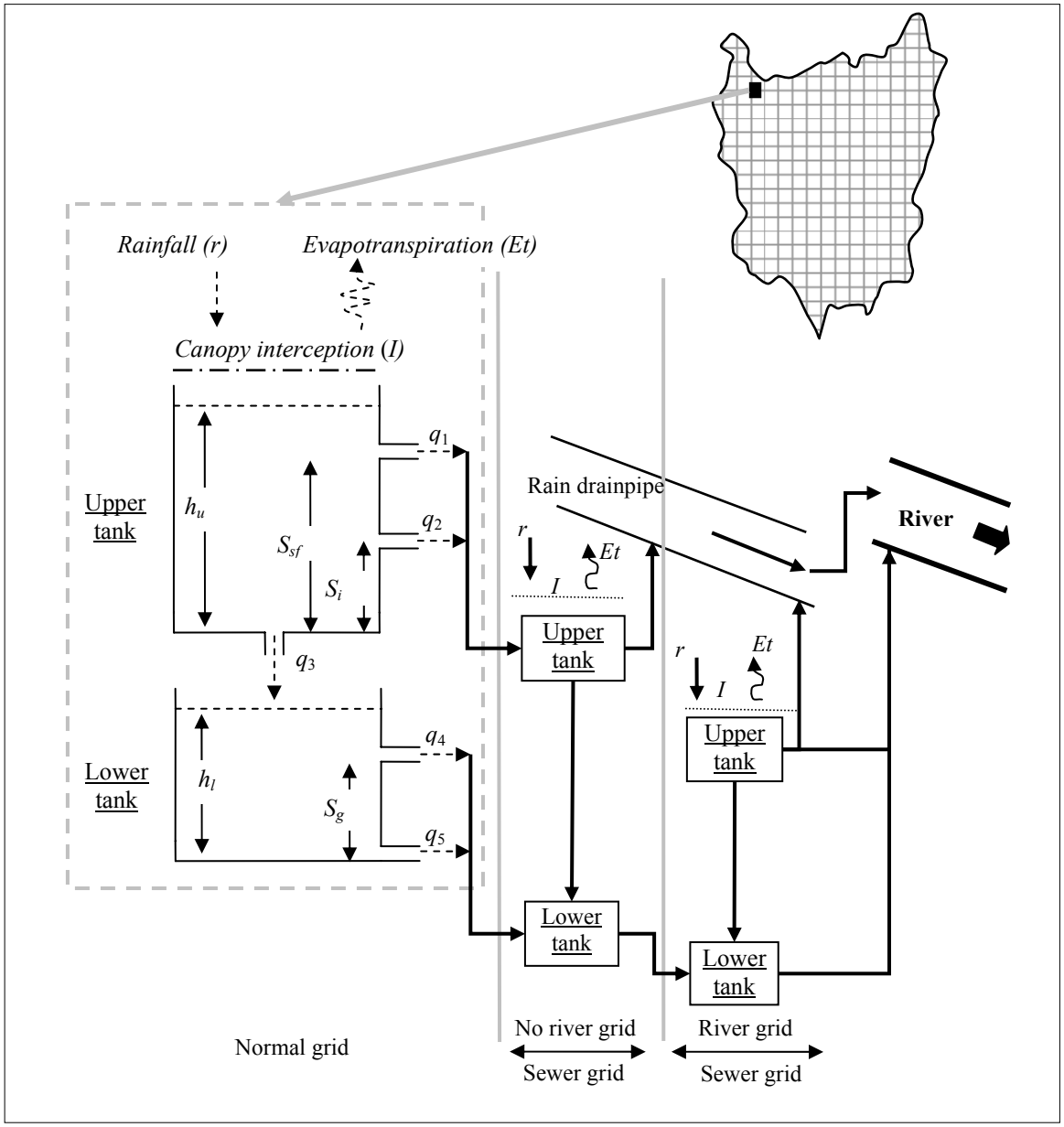


Figure 5.2 Conceptual overview of distributed tank model used in this study

5.4 Development of a distributed water quantity and quality model

The form of distributed water quantity and quality model developed for this study had essentially similar structure as compared with other distributed tank models (Suzuki *et al*., 1996; Xu *et al.*, 2003; Yoshino *et al.*, 1991, Ihara *et al.*, 2003). The model was based on two sub-models, namely, hydrological and water quality sub-models. The details of two sub-models are described hereafter.

5.4.1 Hydrological sub-model

The sub-model conceptualized two storey tanks in each grid. The basic conceptual structure of the sub-model is shown in **Fig. 5.2**. The upper tank was used for the simulation of surface (q_1) and sub-surface (q_2) runoff through the two side outlets. The water from the upper tank percolated (q_3) to the lower tank through the bottom outlet. Besides, upper tank had the function to receive the rainfall, intercept some rainfall by the vegetation canopy, and lose the water through evapotranspiration. Lower tank conceptualized the ground water flow (q_4 and q_5) through two side outlets, but without any bottom outlet. Lower tank was assumed to be at very deep, so that water form it could not go back to the upper tank when the upper tanks were empty (or if soil was dry). In addition, the sub-model identified three types of grids, namely, normal, sewer, and river. In case of normal girds, two-storey tank structure, as described earlier, was conceptualized. In case of sewer grid, the upper tank was transformed into channel without any outlets, while lower tank structure was kept same. For river grids, they were transformed into single channel without any tanks or side outlets. The governing equations related with outputs from each tank were given by the **Equations 5.1 ~ 5.5**.

$$q_1 = (h_u - S_{sf}) \cdot s \cdot f_1 \quad \{h_u > S_{sf} \qquad \textbf{5.1}$$

$$q_2 = (h_u - S_i) \cdot s \cdot f_1 \quad \{h_u > S_i \qquad \textbf{5.2}$$

$$q_3 = h_u \cdot f_0 \qquad \textbf{5.3}$$

$$q_4 = (h_l - S_g) \cdot g_1 \quad \{h_l > S_g \qquad \textbf{5.4}$$

$$q_5 = h_l \cdot g_1 \qquad \textbf{5.5}$$

Where, h_u was the storage water level in the upper tank (or channel), h_l was the storage water level in the lower tank. S_{sf} was the critical outlet height for surface runoff and S_i was the critical outlet height for sub-surface runoff. s was the slope of the grid, f_1 was the runoff factor for the side outlets for upper tank, f_0 was the conductivity rate of bottom outlet of upper tank, S_g was the critical outlet height for groundwater runoff (q_4), and g_1 was the lateral conductivity rate of the outlets of lower tank.

5.4.2 Water quality sub-model

Diffuse pollution is one of the major environmental issues because it could significantly impair the quality of receiving water, such as from eutrophication. Because of

distributed sources of nutrients or pollutants, management of diffuse pollution is more difficult as compared with point sources. In recent days, application of distributed models to estimate the contribution of diffuse sources on water quality was seen as potential option (AnnAGNPS, ANSWER-2000, SWIM, PROW). Distributed models could independently model the variation in each grid, which is one of the basic requirements for the assessment of the impact of diffuse sources on the water quality.

In this study, distributed tank model was integrated with the water quality sub-model based on some fundamental assumptions. Water quality is simulated based on constant buildup and washoff of the water quality indices (termed as pollutants in this chapter) (Zhang and Yamada, 1996; Rossman, 2005; Chen *et al*., 2006; Kim *et al*., 2006). Depending on the types of pollutants, buildup was mainly governed by the natural or artificial accumulation, decay, and removal by physical forces (*e.g*., winds). The equation of pollutant buildup could be written as:

$$w_i = k_0\left(1 - \frac{w_0}{w_{\max}}\right) \qquad \textbf{5.6}$$

Where, w_i (kg/km^2/hr) was the amount of pollutant build up at i^{th} hour, w_0 (kg/km^2) was the amount of unwashed pollutant left, k_0 (kg/km^2/hr) was the constant rate of pollutant buildup, and w_{max} (kg/km^2) was the maximum amount of pollutant that can accumulate per unit area (excluding the transfer by the runoff from upstream areas). Accumulation at the initial stage could occur at the potential rate (k_0) so that it could increase exponentially. However, the accumulation started to decline and eventually attained a plateau when w_i approached w_{max}. It was assumed that with the natural or artificial accumulation of the pollutants over time they were also removed constantly due to several factors (*e.g.,* natural decay, removal by wind forces *etc*.) other than washoff by the runoff. Therefore, when runoff could not occur for long time, especially in the dry season, pollutant accumulation would soon reach the w_{max}.

Washoff was mainly determined as the function of estimated volume of total runoff (mm/hr) from a grid, as given by a power function. The total amount of washoff (L_{out}) was given as follows:

$$L_{out} = (aQ^c) \cdot V \qquad \textbf{5.7}$$

V was the pollutant storage available for washoff (kg/km^2), Q was total volume of runoff (mm/h) from any grid, a was the washoff or sweeping coefficients, and c was the washoff exponent. Although the equation was empirical in nature, a was generally assumed to be related with the properties of each pollutant showing their resistance to the runoff forces. c was related more with the effect of runoff forces on the specific type of pollutants (Morgan, 1995; Asselman, 2000; Xu *et al.*, 2005). Therefore, values of both a and c were specific to the types of pollutants, and are usually determined during the calibration of the model or through the experience from previous studies.

5.4.3 Routing of water and pollutants

Routing of water or pollutants was based on the drainage direction. Rules of drainage direction were described in the Chapter III and Chapter VI. The water and pollutant storage of each tank in all grids were always maintained according to the basic principle of tank model. The volume balance of upper and lower tanks for water quantity were given in **Eq. 5.8** and **5.9** respectively, whereas the mass balance of pollutants storage (V) was given in **Eq. 5.10**.

$$\frac{dh_u}{dt} = q_{input-ut} + r - Et - q_1 - q_2 - q_3 \qquad \textbf{5.8}$$

$$\frac{dh_l}{dt} = q_{input-lt} + q_3 - q_4 - q_5 \qquad \textbf{5.9}$$

$$\frac{dV}{dt} = L_{input} + w - L_{out} \qquad \textbf{5.10}$$

In which, $q_{input\text{-}ut}$, $q_{input\text{-}lt}$, and L_{input} were the amount of water (upper and lower tanks) and pollutants transferred from the neighboring upstream grids (if any) respectively. r was rainfall intensity (mm/hr), Et was the evapotranspiration rate (mm/hr). However, transfer of water or pollutants from neighboring upstream grids to the downstream grid depended on the grid type of latter one (Fig 5.2). For normal grid, water from the neighboring upstream tanks was transferred to the respective tank in the downstream (*e.g.*, upper tanks → upper tank). For sewer grids, the outputs from upper tank were transferred to the channel, but output from lower tank was transferred to the lower tank. In case of river grid, outputs from both upper and lower tanks were transferred to the river channel. Therefore, the output from the lower tank could finally contribute only to the river grid. The outputs

from channel (river or sewer grid) were directly transferred to the interest outlet points depending on the time of concentration, *i.e.*, time required for the water or pollutant to travel a certain distance. However, in this study one hour was used as the time of concentration for all channel grids assuming rapid rainfall-runoff response. To account for the channel runoff, following equation was used for the surface runoff from the river or rain drainage pipes in sewer areas:

$$q_1 = h_u \cdot f_r \qquad \textbf{5.11}$$

Where, f_r (1/hr) was the runoff coefficient for the channel grid. The governing differential equations were numerically solved by the classic fourth order Runge-Kutta method. The basic time step for the simulation was set to be one hour, however during peak runoff periods the time step was slightly reduced to prevent the diversion in the numerical solution. For the flexibility in calculations and processing of data, *MS-Excel* and accompanying *VBA* software were mainly used.

5.5 Values of major parameters

Values of parameters in a distributed model are difficult to adjust, only through the calibration, due to the uniqueness of values for each grid. Therefore, values of parameters of a distributed model should be assigned directly from distributed data sources. Even with the simple structure of the model, values of number of parameters need to be determined for each grid. Rainfall was one of the important parameter that generally showed wider variability in both space and time, but they constituted the major input to the model as shown by Eq. 5.8. Daily rainfall data for each grid was already determined in Chapter III, but due to the hourly time step of the simulation the daily rainfall intensity (mm/d) needed to be transformed into hourly scale (mm/hr). The daily rainfall data observed at the main weather stations (W) ($r_{main\text{-}daily}$) was highly correlated (r: 0.86-0.99) with other rainfall stations (R1 ~ R5) data ($r_{other\text{-}daily}$) inside the watershed, but W also had the rainfall data observed every hour ($r_{main\text{-}hourly}$). Therefore, using available rainfall data ($r_{other\text{-}daily}$, $r_{main\text{-}daily}$ and $r_{main\text{-}hourly}$), hourly scale rainfall ($r_{other\text{-}hourly}$) for each grid was created as shown by the following equations:

$$y_i = \frac{(r_{main-hourly})_i}{r_{main-daily}} \qquad \{i = 1,2,3.....24 \qquad \textbf{5.12}$$

$$(r_{other-hourly})_i = y_i \times r_{other-daily} \qquad \{r_{main-daily} > 0;\ \ i = 1,2,3.....24 \qquad \textbf{5.13}$$

$$(r_{other-hourly})_i = \frac{r_{other-daily}}{24} \qquad \{r_{main-daily} = 0;\ \ i = 1,2,3,.....24 \qquad \textbf{5.14}$$

Canopy interception (*I*) and evapotranspiration (*Et*) were determined from the processing of satellite images and *FAO* Penman-Monteith method respectively (details will be explained in Chapter VI).

For other parameters, mainly land cover, geology, soil profile depth, and topographic variables were used to define the values of major parameters. In hilly areas, influence of topography could have dominant influence on both surface and subsurface runoff due to the accelerative effects of gravity. Slope was one of the important measures of topography, which was often used directly in the estimation of runoff from the surface (*e.g.*, Manning's equation) or sub-surface (*e.g.*, Darcy's equation for the subsurface runoff in hilly area). Therefore, runoff from the outlets (q_1 and q_2) from the upper tank was assumed directly proportional to the slope. Besides, it was also assumed that the land cover of each grid could affect the surface runoff due to the variations in the surface roughness, conductivity of subsurface, or permeability of the bottom outlet. Therefore, a runoff coefficient (f_1) (hr^{-1}) was introduced in the equations for both q_1 and q_2 (Eq. 5.1 and 5.2), while the percolation factor (f_0) (hr^{-1}) was assumed linearly proportional to the water level (h_u) q_3 (Eq. 5.3). The values of coefficients for major land cover were adjusted to appropriate values during the model assessment process.

Critical outlet heights (S_{sf}, S_i, and S_g) were quite important for the generation of surface, subsurface and ground water runoff, and the criteria to assign its values were not well defined. Three main watershed variables could be related with S_{sf} and S_i, namely, land cover types, depth of soil, and topography. We used soil topographic index (Beven and Krikby, 1979) of TOPMODEL to determine the value of S_{sf} and S_i. The concepts of topographic index in TOPMODEL and distributed tank model are depicted in **Fig. 5.3**. TOPMODEL was a semi-distributed model based on the variable source contributing area concept, which states that overland flow was produced only from the certain fraction of the watershed that attains early saturation during the rainfall.

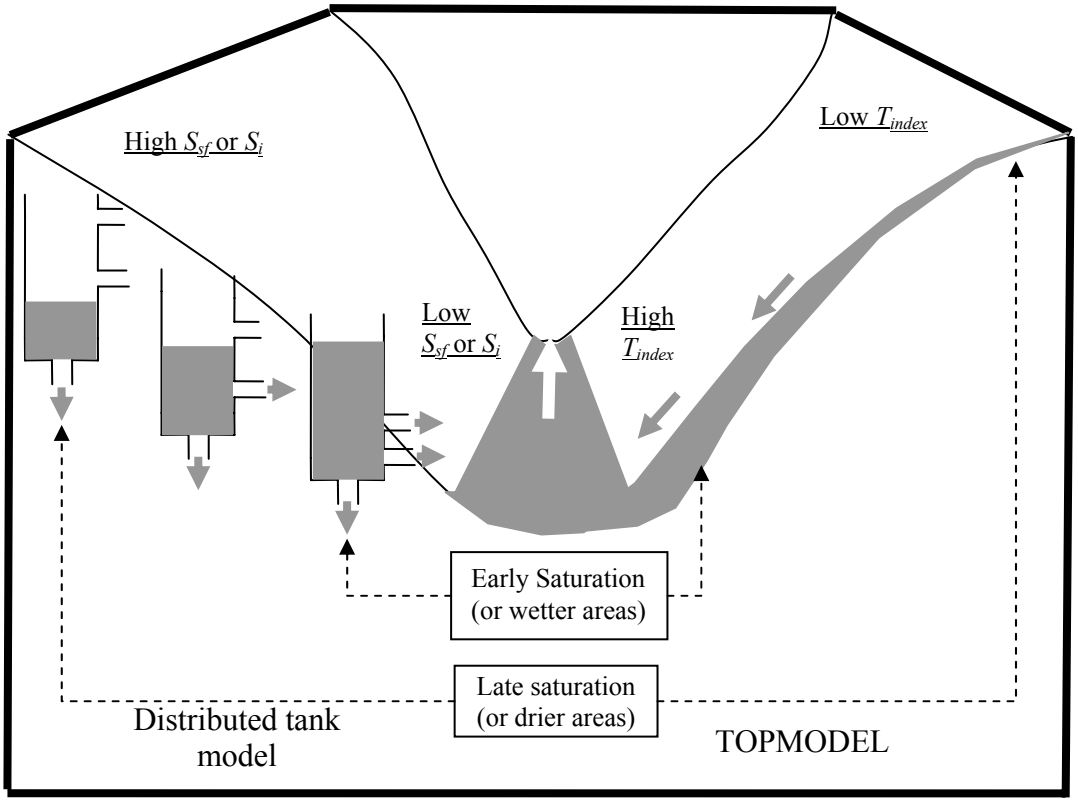


Figure 5.3 Overland runoff generation in a hilly terrain on the basis of soil topographic index distribution conceptualized in distributed tank model and TOPMODEL.

The main criteria to classify variable source contribution areas were determined by the soil topographic index value at a certain point of the watershed. The soil topographic index (T_{index}) could be written as:

$$T_{index} = \ln\left(\frac{a}{T_0 \cdot \tan\beta}\right) \qquad \textbf{5.15}$$

Where, a was the upslope contributing area per unit grid length (m^2/m), $\tan\beta$ was the slope of the grid, and T_0 was the average transmisivity (m^2/s). The places having higher value of a and lower value of $\tan\beta$ (or T_0) had higher possibility of earlier saturation and vice versa. So the areas having similar topographic index value were assumed to show similar hydrological behavior. The concept was not only relevant to the TOPMODEL but also formed the basis of modeling in hill slope hydrology (Beven, 2001). In this study, we assumed that grids having similar value of soil topographic index should have similar value of S_{sf} or S_i (Fig 5.3). Land cover was not considered since majority of the study area had

forest cover. In **Eq. 5.15** it could be realized that $T_0 = \overline{K} \times D$, where $\overline{K}$ was the average hydraulic conductivity (m/s); and D was the average depth of the soil profile (m). Assuming $\overline{K} = 1$ for whole watershed, the index reduces to:

$$T_{index} = \ln(\frac{a}{D \cdot \tan \beta}) \quad \textbf{5.16}$$

D was available from the geology map, but all other parameters of **Eq .5.16** were derived from the analysis of DEM. Then critical heights for surface and subsurface runoff were determined as:

$$S_{sf} = \frac{Z}{T_{index}} \quad \textbf{5.17}$$

$$S_i = \frac{S_{sf}}{4} \quad \textbf{5.18}$$

The value of Z was chosen 90 mm to derive the S_{sf} into a satisfactory range of 10 – 45 mm, *i.e.*, at least 10 - 45 mm of rainfall was required for the surface runoff depending on the T_{index} for a given grid. S_i was assumed 1/4th of S_{sf}, but it could be changed to more suitable one. Both **Eq. 17** and **18** stated that grids with lower values of T_{index} had lesser possibility of surface or subsurface runoff (generally at the top of the hill), but grids with higher values of T_{index} were more likely to generate the surface and sub-surface runoff. The value of S_g was assumed constant for all grids assuming that the lower tank could have only marginal effect on the river runoff, which could mainly contribute the baseflow rate. The g_1 was based on the permeability of underground geology, which was adopted from another study in Japan (Suzuki *et al.,* 1996).

The production of the pollutants (k_0) in each grid was based on the land cover type and population (capita/grid). However, population loadings in the sewer grids were ignored because they were normally transferred outside of the basin. Values of other parameters were adopted from literature, but they were later on modified in the model assessment process to fit with the observed data.

5.6 Overview of Microsoft (MS) Excel as graphical user interface (GUI) used in this study

MS-Excel is a software distributed by *Microsoft Inc*. as a component of office programs. It is powerful software to deal with data, do calculations, and to perform various analyses. Besides, it also provides an advanced programming environment, Visual Basics for Application (*VBA*), which could significantly enhance the capability of the software even to perform the range of calculations, such as in modeling. A good example of the application of *MS-Excel* in river water quality modeling was the QUAL2K model of USEPA that also utilized *VBA* environment to perform calculations and visualize the outputs in charts or diagrams (Chapra and Pelletier, 2003).

In this study, we mainly used *MS-Excel* to perform all of the calculations in the modeling as well as a GIS environment. In a *MS-Excel* workbook, there are worksheets, each of which is characterized by large number of fundamental units or cells arranged in a matrix form. Each cell therefore could be identified from their respective row and column number. If a cell was assumed as a raster grid, *MS-Excel* could be thought as a GIS environment. We assumed a worksheet as raster layer, and performed the raster processing by utilizing the *VBA* environment. If special processing of raster data was needed, the worksheet could be exported to generic ASCII format, which was easily accepted by GIS software, such as *ArcGIS* 8.3. In the model, all grids were assigned their unique identity (ID) on the basis of drainage direction, so that all values of model parameter for each grid could be identified with that ID. Using the ID for each grid, various modifications of the attributes as well as routing of the water and pollutants could be done quite effectively.

5.7 Summary

With increasing demands on water resources and emergence of several water related problems, tools for improved decision-making would be required for the effective water resources management. Especially, distributed models could be the only practical options to deal with range of water resources problems, such as diffuse pollution, in future. Due to the limitations in the understanding of non-linear hydrological processes as well as on the analytical methods, some uncertainty in estimated outputs always exists even by using the conceptually perfect and physically sound distributed models. Besides, our limited capability to measure the hydrological as well as chemical processes of larger spatial

interest often complicates the uncertainty of the distributed models. Therefore, simple and less parameter intensive distributed model that have the potential to represent governing hydrological as well as water quality variations are required for practical application.

Recently, remote sensing and GIS have proven as efficient techniques to deal with larger spatial data required by the distributed models, and may contribute to minimize uncertainty of the estimated values. This study aimed to develop a simple distributed water quality model capable of using both remotely sensed data and GIS environment. Following were the important points emphasized in this chapter:

a) A distributed model based on the principle of lumped tank model was developed, which was characterized by its simple conceptual framework, less parameter intensive, and familiar *MS-Excel* based graphical user interface for flexibility in re-design and application;

b) Major input parameters of the model were based mainly on the remotely sensed data and digital elevation model (DEM), which was a significant improvement over the lumped type tank model that mainly rely on calibration and validation for the values of its parameters;

c) The model was expected to be useful to estimate the river discharge and river water quality as an output from diffuse sources, especially, from the small, hilly and forested watersheds

CHAPTER VI

APPLICATION OF REMOTE SENSING AND GEOGRAPHIC INFORMATION SYSTEM (GIS) IN THE DISTRIBUTED MODEL

6.1 Introduction

Remote sensing and GIS are powerful techniques to deal with the spatial data. Remote sensing technique is mainly focused on generating the spatial data by analyzing the response of electromagnetic radiation on different environmental variables. GIS is an efficient technique that can store, process, display, and analyze the spatial data in order to yield meaningful spatial variables. Both remote sensing and GIS are essential to distributed modeling since the model require a lot of spatial data to assign the values of its parameters. Besides, as described in Chapter III, processing of the spatial data are necessary to create essential features, such as drainage direction for routing of water and pollutants and to organize the spatial data for each gird that could be easily assessed during the simulation.

This chapter was mainly focused to generate the important spatial data required by the distributed model by the use of remote sensing. Besides, digital elevation model (DEM) was also processed with the help of GIS to create meaningful topographic variables required by the model during parameterization.

6.2 Methodology

Application of remote sensing was mainly focused to create land cover data, leaf area index (*Lai*), and transpiration coefficient (T_c). Land cover data was used to assign several model parameters (such as outlet factors and pollution production). *Lai* was utilized to determine the intercepted rainfall by the forest canopy. T_c was used as an input for the estimation of evapotranspiration loss. GIS was mainly used to process the watershed attributes and organize them in the model. Since elevation has significant role in the hydrological processes, major portion of GIS application was focused in the processing of DEM to derive important topographic variables, such as, slope, drainage direction, upslope

contributing area, shaded relief, and topographic index. Processing of all satellite images and derivation of shaded relief was done by *ENVI* 4.2 software (*Research Inc.*). GIS processing was done mainly by the use of *MS-Excel*-2003 and accompanying *VBA* (*Microsoft Inc.*), but *ArcGIS* 8.3 (*ESRI Inc.*) was also used in some cases.

6.2.1 Land cover classification

Land cover or land use data were undoubtedly the most widely used spatial information in many disciplines and it is the most widely used input in distributed modeling. Applications of remote sensing were mainly focused on the classification of land cover or derivation of land use maps since the historic time. Land cover mapping was the major potential of remote sensing technique (Jensen, 1983) and it could be done directly with the help of conventional classification techniques in many instances (Mather, 1999; Lillesand and Kiefer, 2000). However, for the specific cases, it could be significantly complex, and even until now, remote sensing researches are focused mainly on the land cover classification (*e.g.*, Dymond and Shepherd, 2004; Evans and Geerken, 2006).

In our study, major use of land cover map was to identify the modeling parameters related with water quantity and water quality. In the water quantity sub-model, land cover data were required to estimate the potential of canopy to intercept the rainfall, evapotranspiration losses, and to assign the runoff parameters of the outlets. In case of water quality sub-model, land cover data were the only basis to assign the loadings parameters since the loadings from an area were usually linked with its land cover types.

Among the available classification techniques, we used a hierarchical binary decision tree classification (Friedl and Brodley, 1997; Mclver and Friedl, 2002; Pal and Mather, 2003; Im and Jensen, 2005). Decision tree is a non-parametric classification technique of remotely sensed data, in which statistical distribution of the data is not important thereby providing flexibility during the classification process (Friedl and Brodley, 1997). Another major benefit of decision tree classification was that several data sources, other than the remote sensing inputs, could be utilized to enhance the classification process.

In a binary decision tree, there are nodes and decision branches as shown in the **Fig. 6.1**. In each node, one or more criteria could be defined that would give binary results (or two branches). Again new criteria could be developed for each decision branch, which transforms the branch into a new node, yielding two more branches from that. In that way,

the process could be continued until the desired level of classification was reached. Initial mode of classification in decision tree are generally targeted to distinguish the general classes, but proceeding further could yield more specific classes making the decision tree more complex.

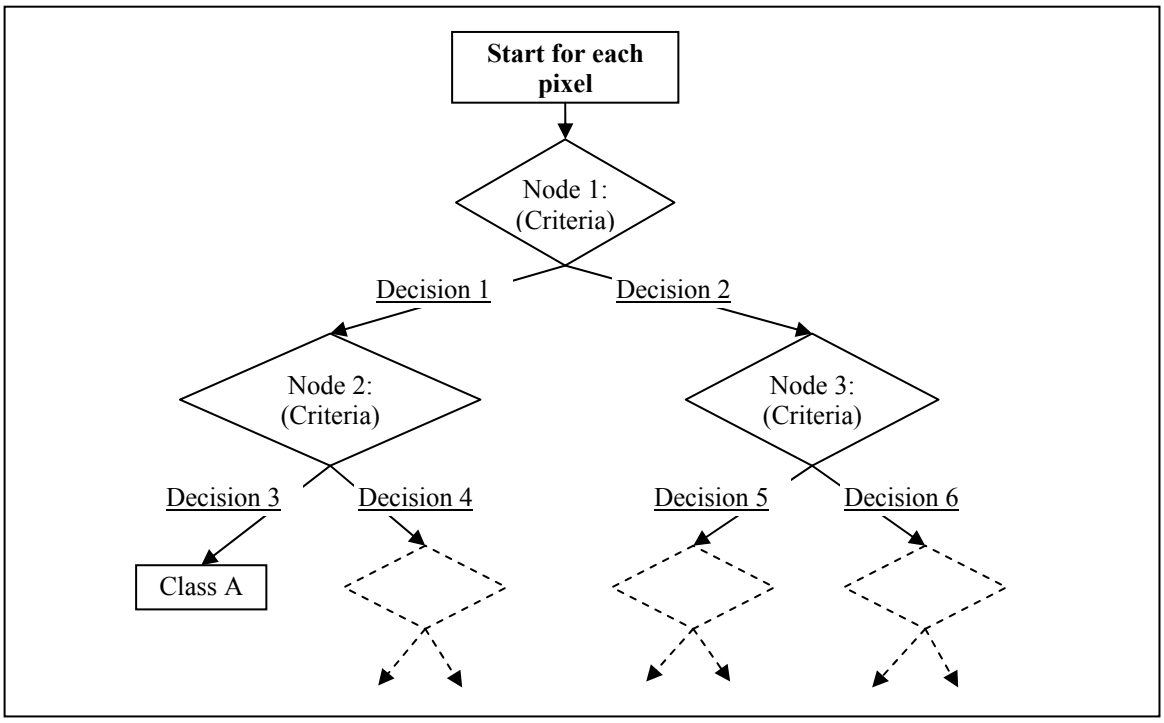


Figure 6.1 A binary decision tree classifier

In our study, we mainly used the multiple date ETM+ images and topographic variables to classify the land cover of the area. Since majority of the study area had forest cover, so major emphasis was given to identify major forest types, such as deciduous, evergreen or mixed. Because of the uniqueness of each watershed, no single criteria could give same results at different places. Therefore, we mainly focused on using spectral patterns of key features specific to the study area. In addition, common indices (*Ndvi*, *Ndwi*), visual analysis of each band, and topographic variables (slope and shaded relief) were used to develop the decision tree. In which, shaded relief was calculated from the DEM for each ETM+ image by using respective image azimuth and sun elevation (Appendix A). **Figure 6.2** shows the process used while designing the decision tree in this study. Initially visual analysis of study area was performed with the help of each bands and common indices (*Ndvi* and *Ndwi*) to identify key features. The common indices could be written as:

$$Ndvi = \frac{TM4 - TM3}{TM4 + TM3} \quad \textbf{6.1}$$

Where, *Ndvi* was the normalized differential vegetation index. *TM*3 and *TM*4 were the red and near-infrared bands respectively, in which former showed strong absorption and the latter one showed strong reflectance with *chlorophyll a* in the green leaves of vegetation. Higher *Ndvi* indicates the maximum vegetation cover and lower value indicates few or without significant vegetation. Similarly, normalized differential water index (*Ndwi*) could be written as:

$$Ndwi = \frac{TM4 - TM5}{TM4 + TM5} \quad \textbf{6.2}$$

Where, *TM*5 was the mid-infrared band. Since *TM*5 was sensitive to the moisture content, *Ndwi* was found to be very effective to detect the moisture changes in the vegetation in comparison with *Ndvi* (Wilson and Sader, 2002). In our study area, deciduous species undergo significant physiological leaf color change during winter in comparison with evergreen species, so we used *Ndwi* to detect the physiological changes. During winter the moisture content in the leaves of deciduous vegetation were supposed to be lower than evergreen species. In such conditions, *TM*5 would show higher reflectance, but *TM*4 may not show significant differences between green and non-green vegetations. Therefore, higher *Ndwi* would show more green species than the lower *Ndwi* values. Major endmember types (spectral pattern) of key features were then extracted with the help of endmember extraction tool available in the *ENVI* 4.2. The major endmembers were then compared to derive a new index (*Ni*) as:

$$Ni = TM4 - TM5 - TM7 \quad \textbf{6.3}$$

Where, *TM*7 was the far-infrared band, which was also found sensitive to the moisture content of the vegetation in this study.

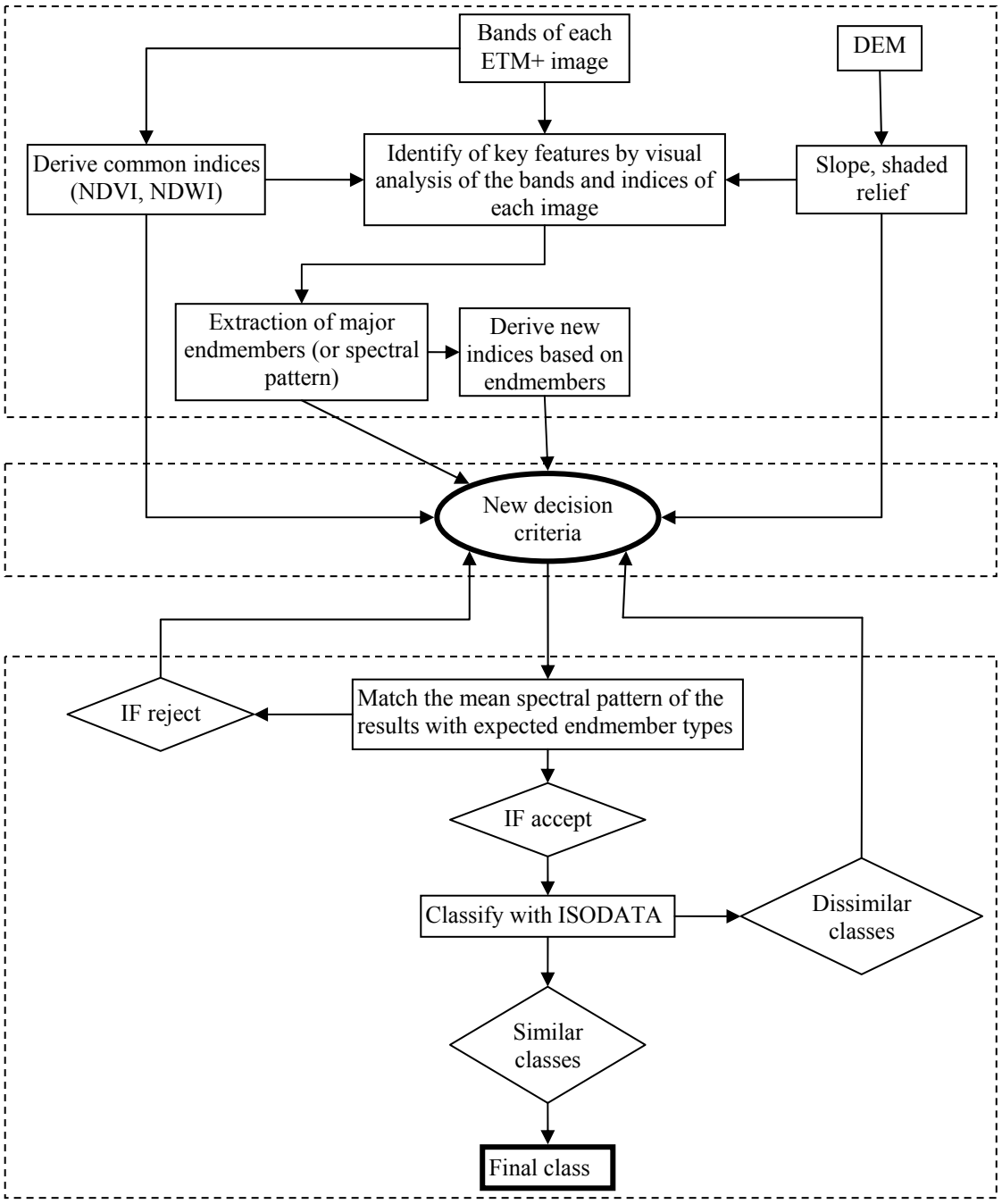


Figure 6.2 Flow chart showing the development of decision tree classification scheme used in this study

Based on above information decision criteria were developed. Decision criteria could be affected by subjective judgments, so their reliability needs to be confirmed. For that purpose, results from each decision were compared with the expected endmember types. If the results were found convincing, the classified portion were subject to iterative self-organizing data analysis (ISODATA) unsupervised classification. The classes from the ISODATA were then combined on the basis of similarity of their mean endmembers to

yield the final land cover class. However, the dissimilar classes (if any) were again subject to the new decision criteria. The results were further assessed with the help of air photograph, and with the help of pictures taken at some points inside the watershed (**Appendix C-1**).

6.2.2 Leaf area index (*Lai*) and canopy interception (*I*) capacity

Lai is one of the important variables, which represents the vegetation density in an area. In this study, *Lai* was mainly utilized to estimate the rainfall interception by the canopy. *Lai* could be defined as the total area of leaves (one sided) in a unit area of land. *Lai* is therefore unit less (m^2/m^2) and the values greater than one represents the completely vegetated area whereas less than one represents areas with sparse vegetation or no vegetation at all. Various possibilities of applying remote sensing technique to estimate the *Lai* from the satellite images had been widely assessed. In most cases, wavelength corresponding to the *TM*3 and *TM*4 were used to estimate *Lai* due to the unique spectral response of both bands to the vegetation. It was due to the close relationship between these two bands with the photosynthetically active radiation (*PAR*) absorbed by the green vegetations. In our study, we used image of October to derive the green *Lai* because in that period all of the vegetations were supposed show green.

There were several methods to determine the *Lai* ranging from empirical methods to theoretical ones. We used a method in Simple Biosphere Model 2 (SiB2) to determine the *Lai* of major vegetation types in the study area (Sellers, 1996a). SiB2 was an improvement of in SiB1 method, which aims to calculate turbulent transfer and reflectance properties of vegetation related land surface as functions of *Lai*, canopy morphology, and vegetation element, and soil optical properties. In SiB2 method, improvements were mainly done on the part of realistic canopy photosynthesis-conductance sub-model, and the use of satellite data to describe vegetation state and phenology (Sellers *et al.*, 1996a). Details could be assessed from the relevant references (Sellers *et al.*, 1996a,b), however, use of satellite data to estimate fraction of photosynthetically active radiation absorbed by the green vegetation canopy (*Fpar*) and *Lai* are discussed here(Sellers *et al.*, 1996b). The estimation process had two steps:

a) Calculation of *Fpar* from simple ratio (or ratio vegetation index) (*Sr*):

Simple ratio could be calculated directly from the *TM*3 and *TM*4 bands of ETM+ or from

Ndvi image:

$$Sr = \frac{TM4}{TM3} = \frac{1 + Ndvi}{1 - Ndvi} \qquad \textbf{6.4}$$

Assuming linear relationship, an equation between *Fpar* and *Sr* could be established if *Ndvi* at maximum and minimum *Fpar* were known (Sellers *et al.*, 1996b). For the convenience, the maximum and minimum could be assumed at 98% population of *Ndvi* of each vegetation type, when *Fpar* was nearly one, and at 5% *Ndvi* population nearly resembling to the desert area without any vegetation, when the *Fpar* was nearly 0.001. Then relationship between *Fpar* and *Sr* could be given as:

$$Fpar_i = Fpar_{\min} + (Fpar_{\max} - Fpar_{\min}) \frac{(Sr_i - Sr_{\min})}{(Sr_{\max} - Sr_{\min})} \qquad \textbf{6.5}$$

Where,

$Fpar_i$=*Fpar* of i^{th} grid;

$Fpar_{\min}$=0.001 (independent of land cover) (Zhou *et al*., 2006);

$Fpar_{\max}$=0.980 (independent of land cover) (Zhou *et al*., 2006);

Sr_i = *Sr* of i^{th} grid;

$Sr_{\min}$ = *Sr* value corresponding to 5% of *Ndvi* population of a vegetation type;

$Sr_{\max}$ = *Sr* value corresponding to 98% of *Ndvi* population of a vegetation type, which was set 0.039 as global value (Zhou *et al*., 2006);

b) Calculate green *Lai* from *Fpar*:

The relationship between *Fpar* and the green *Lai* could be written as:

$$Lai_i = Lai_{\max} \frac{\ln(1 - Fpar_i)}{\ln(1 - Fpar_{\max})} \qquad \textbf{6.6}$$

Where,

Lai_i = *Lai* for i^{th} grid;

$Lai_{\max}$ = Maximum *Lai* for a given vegetation type;

The relation was assumed applicable for evenly distributed vegetations, but for the clumped vegetations, such as coniferous trees, the equation becomes:

$$Lai_i = Lai_{\max} \frac{Fpar_i}{Fpar_{\max}} \quad \textbf{6.7}$$

However, to be applicable for all vegetation conditions, the equation transforms into the following form:

$$Lai_i = (1 - F_{cl}) Lai_{\max} \frac{\ln(1 - Fpar_i)}{\ln(1 - Fpar_{\max})} + F_{cl} Lai_{\max} \frac{Fpar_i}{Fpar_{\max}} \quad \textbf{6.8}$$

Where, F_{cl} was the fraction of clumped vegetation. The recommended values of unknown parameters were used from Zhou *et al*. (2006) as shown in **Table 6.1**.

Table 6.1 Recommended values of parameters used to determine *Lai*

SN	Vegetation type	$Lai_{\max}$	F_{cl}	$NDVI_{98\%}$
1	Urban vegetation	6.5	0.5	0.674
2	Shaded vegetation	5.7	0.5	0.721
3	Decidious forests	5.5	0	0.721
4	Mixed forests	5.7	0.5	0.721
5	Evergreen forests	7	1	0.689

For the estimation of canopy interception (*I*), we used the following relationship:

$$I_{i,\max} = C_{\text{int}} Lai_i \quad \textbf{6.9}$$

Where,

$I_{i,\max}$ = Maximum interception capacity of i^{th} grid;

C_{int} = Interception coefficient (mm);

We used 0.2 mm as the value of C_{int} (Zhou *et al*. 2006). Interception balance of a grid was then formulated according to the following assumptions:

a) A grid could intercept the rainfall not more than $I_{i,\max}$;

b) All intercepted rainfall was lost through evaporation;

c) Interception capacity of deciduous vegetation were zero during dry period (December – April), when leaves were supposed to start falling on the ground.

6.2.3 Evapotranspiration (*ET*) estimation

Evapotranspiration (*Et*) is a physical process, in which water is lost to the atmosphere from evaporation and through transpiration from the living plants. *Et* loss is very important in modeling since it could significantly affect the water balance of the watershed. Vegetation had dominant contribution on the *Et* loss, and *Et* rate largely depends on the vegetation types, in addition to other biophysical factors. *Et* from a given grid area could be written as:

$$Et_i = T_{i,c} Et_0 \qquad \textbf{6.10}$$

Where,

Et_i was the evapotranspiration rate from i^{th} grid (mm/d);

$T_{i,c}$ was the transpiration coefficient of a i^{th} grid;

Et_0 was the reference evapotranspiration rate (mm/d);

There were several methods to calculate Et_0, such as observation from Pan Evaporation, Penman method, or Priestley-Taylor. Recommended values of T_c were only available for certain crops (as crop coefficient) (Allen *et al*., 1998), but for forest or other vegetations the values could not be found easily. As an alternative, spectral reflectance of remotely sensed data could provide an indirect method to calculate the values of T_c. In this study, we used a scaling method proposed by Choudhury *et al*. (1994) to estimate the values of T_c. Where, T_c was defined as the ratio of unstressed *Et* and Et_0. It should be noted that when crop coefficient is used it is the ratio of total *Et* (irrespective of the water availability) and Et_0. Choudhury used a heat balance and radiative transfer model to study the relationship between T_c and vegetation indices. Heat balance model solved the equations defining the conservation of heat and mass (water vapor) in the soil-vegetation system. The radiative transfer model calculated the reflectance of red (0.6-0.7 μm) and near-infrared (0.8-1.0 μm) regions, which could be used to calculated different vegetation

indices. From the experiment, Choudhury *et al*. (1994) obtained the direct relationship between T_c and scaled vegetation index, which was given as:

$$T_{i,c} = 1 - \left[\frac{Vi_{\max} - Vi_i}{Vi_{\max} - Vi_{\min}} \right]^n \quad \textbf{6.11}$$

Where, $Vi_{\max}$ and $Vi_{\min}$ were the maximum and minimum value of *Vi* obtained when the area did not have any vegetation ($Lai \approx 0$) and had maximum vegetation ($Lai \approx$ max) respectively. Vi_i was the value of *Vi* for i^{th} grid and $n = (\kappa/\kappa')$. κ was the damping coefficient which had value in the range of 0.5-0.7, and κ' was the damping coefficient specific to the *Vi* used. According to Choudhury *et al*. (1994), the values of κ' appeared in the range of 0.5 - 0.7 if the *Vi* were soil adjusted vegetation index (*Savi*), which could be written as:

$$Savi = \frac{1.5(TM4 - TM3)}{(TM4 + TM3 + 0.5)} \quad \textbf{6.12}$$

Savi was a modification of *Ndvi* that could normalize the effect of soil reflectance on *TM*3 and *TM*4 space allowing the detection of vegetation more effectively (Huete, 1988). In such condition $n \approx 1$, so that **Eq. 6.11** transforms as:

$$T_{i,c} = \frac{Savi_i - Savi_{\min}}{Savi_{\max} - Savi_{\min}} \quad \textbf{6.13}$$

Both of the Eq. 6.1 and 6.13 actually transform the *Vi* into scaled value of 0-1, which represents the relative density of vegetations in a given grid. Although this relation was not tested for forest vegetation, we used it in our study due to the theoretical basis behind its derivation.

For the estimation of Et_0, we used FAO Penman-Monteith method, which was quite popular and widely used in many fields and probably the most physically soundest among the available methods (Allen *et al*., 1998; Zhou *et al*., 2006). According to FAO Penman-Monteith method, "Et_0 could be defined as the evapotranspiration loss from a reference

surface that is covered with a hypothetical crop of 0.12 m height (similar to grass), have a fixed surface resistance of 70 s m^{-1} and has albedo of 0.23". Under such conditions, Et_0 (mm/d) could be calculated as:

$$Et_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad \textbf{6.14}$$

Where,

R_n = net radiation at the hypothetical crop surface (mm/d)

G = soil heat flux density (MJ/m^2.d)

T = mean daily air temperature at 2 m height (^{0}C)

u_2 = wind speed at 2 m height (m s^{-1})

e_s = saturation vapour pressure (kPa)

e_a = actual vapour pressure (kPa)

Δ = slope of vapour pressure curve (kPa/^{0}C)

γ = psychrometric constant (kPa/^{0}C)

The only factors affecting Et_0 were climatic parameters, so it expresses the evaporating power of the atmosphere at a specific location and time of the year. **Eq. 6.14** is recommended because it closely approximates *Et* from a grass, it is physically based, and it explicitly incorporates both physiological and aerodynamic parameters. Details of solving the **Eq. 6.14**, depending on the availability of the data, were well explained in Allen *et al.* (1998). In our study daily weather data, such as temperature (maximum, minimum and mean), mean relative humidity, wind speed, sunshine duration, which was observed at the main station (W) (Fig 3.1) outside of the watershed, were mainly used for the calculation of Et_0. It was well established that temperature usually decrease with increasing elevation, so we used the lapse rate of 0.6 ^{0}C per 100m rise in elevation to adjust the observed temperature for each grid (Xu *et al.,* 2003) during the calculation of Et_0 for each grid:

$$T_{i,adj} = T_{obs}\left[1 - \frac{0.6(h_i - h_s)}{100}\right] \quad \textbf{6.15}$$

$T_{i,adj}$ and T_{obs} were the adjusted and observed temperature at i^{th} grid and weather station respectively;

h_i and h_s were the elevation (m) of i^{th} grid and weather station respectively.

Since the model was run at hourly scale but Et_0 was calculated at daily scale, Et_0 was divided by 24 to convert into hourly scale. Evapotranspiration loss from each grid was calculated depending on the state of canopy storage and vegetation as:

a) When intercepted rainfall by canopy was significant ($> 0.5I_{max}$), then $Et = Et_0$ (mm/hr);

b) When intercepted rainfall by canopy was insignificant ($< 0.5I_{max}$), then $Et = T_c Et_0$;

c) When there was no vegetation ($Lai < 1$), $Et = Et_0$ (mm/hr); and

d) For deciduous vegetation, during dry period (December –April), Et occurred at Et_0 (mm/hr).

Because of frequent rainfall, it was assumed that watershed might remain moist for most of the times, so that in non-vegetated areas Et could occur at Et_0. In vegetated areas ($Lai > 1$), Et could be significantly affected due to the surface resistance of the canopy so consideration of T_c was quite important.

6.2.4 Derivation of slope, drainage direction, and tank height for sub-surface runoff (S_{sf}, S_i)

Slope, drainage direction, upslope contributing area, and topographic index were the main topographic variables derived from the DEM. **Figure 6.3** shows the detailed flow chart of DEM processing to derive different topographic variables used in this study. Slope (s_i) was calculated by steepest descent method as shown by Eq. (6.16)

$$s_i = \frac{Ele_i - Ele_l}{dist} \qquad \textbf{6.16}$$

Where,

s_i was the slope of i^{th} grid;

Ele_i was the elevation of i^{th} grid;

Ele_l was the lowest elevation (most steepest) grid among the eight neighboring grids, and

dist was the distance between two grids (including the diagonal direction).

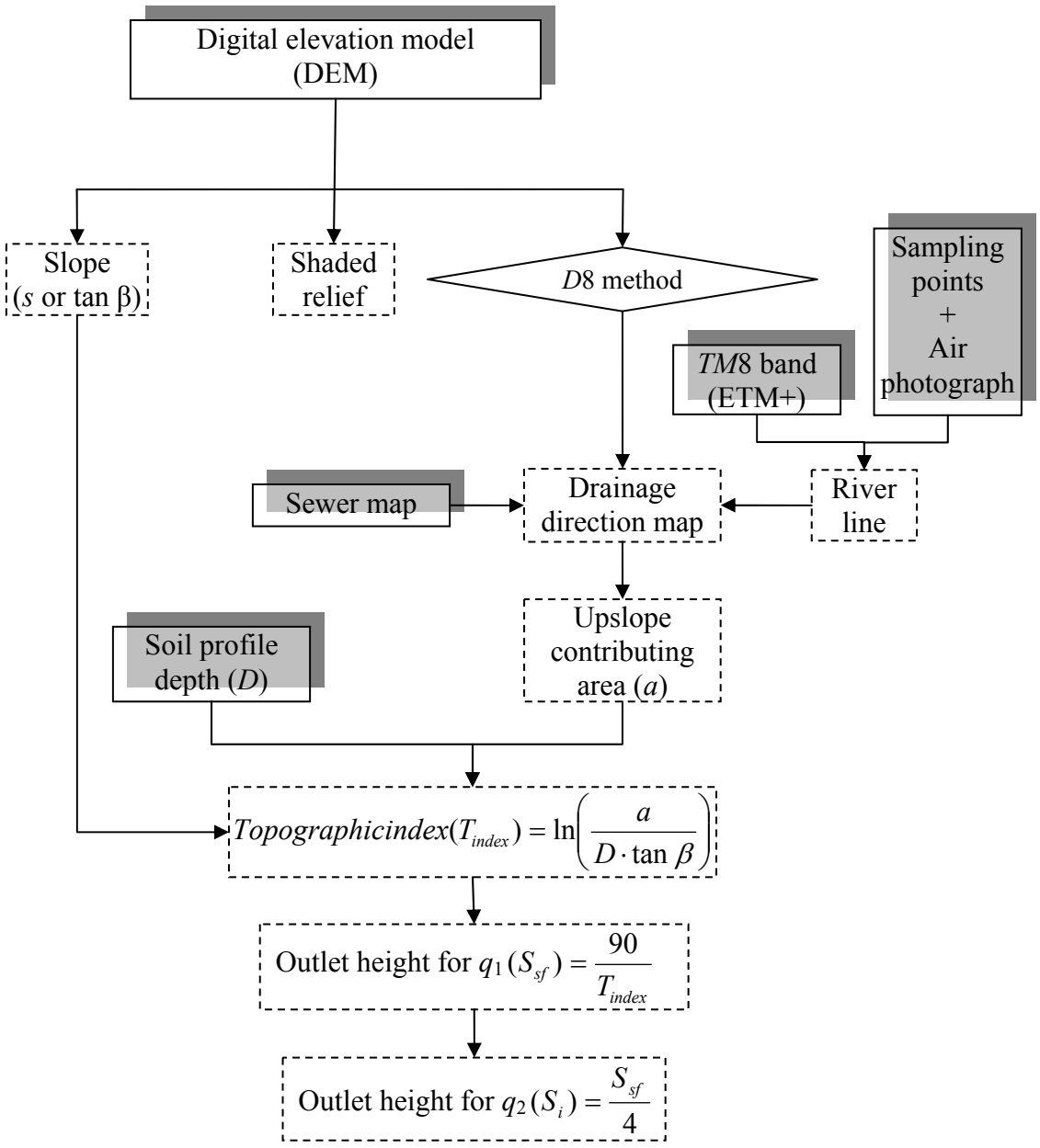


Figure 6.3 Flow chart of processing of DEM and other variables to determine different topographic variables used directly or indirectly in the model.

As mentioned earlier in Chapter III, drainage direction was the only basis to route the water and pollutants from upstream grids to the downstream grids in distributed models. Besides, drainage direction map was required to determine watershed boundary and upslope contributing area, which had great significance to model estimation and determine the saturation excess respectively.

Drainage direction was determined by the D8 method, in which, water could flow towards only one grid, *i.e.*, the steepest, among the eight neighboring grids as shown in the **Fig. 6.4**. However, due to the differences between the scale of DEM and real topography, drainage direction determined by this method sometimes could not coincide with the actual river line. Sewer grids were already available in our database as mentioned in Chapter III, but river line need to be constructed. For that purpose, we used higher resolution panchromatic *TM*8 band of ETM+ image of October. *TM*8 was in grey scale, but its bandwidth was large enough (0.5 - 0.9 μm) to cover the visible range. Therefore, we derived a true color image having similar spatial resolution of *TM*8 (14.25 m) by combining *TM*8 and low-resolution visible bands by the Gram-Schmidt pansharpening technique available in *ENVI* 4.2. Then a river vector line was drawn over the final image with the help of coordinates of some sampling stations, air photograph of the study area, and river line detected by D8 method, in which river line from the most upstream stations was extended until it was easily detectable in the image. Overall process of drawing river line is depicted in **Fig. 6.5**.

We calculated the upslope contributing area of each grid by counting number of upstream grids with the help of drainage direction map. After that soil topographic index was determined as given in the Eq. 5.16. As depth of soil profile were available in three

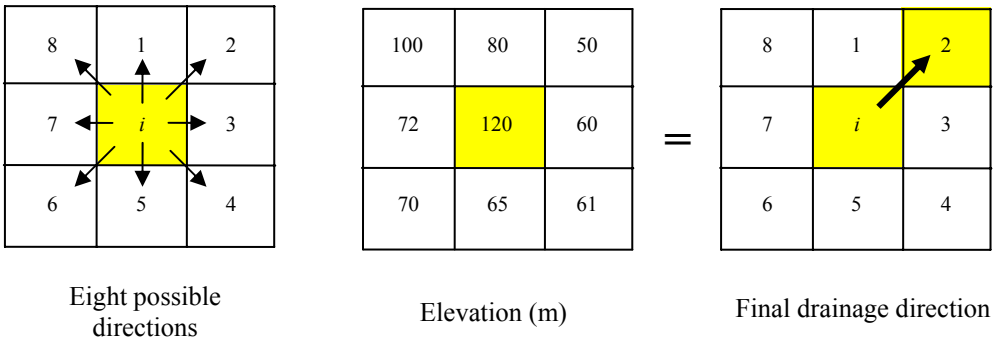


Figure 6.4 Concept showing the method to determine the drainage direction by the D8 method

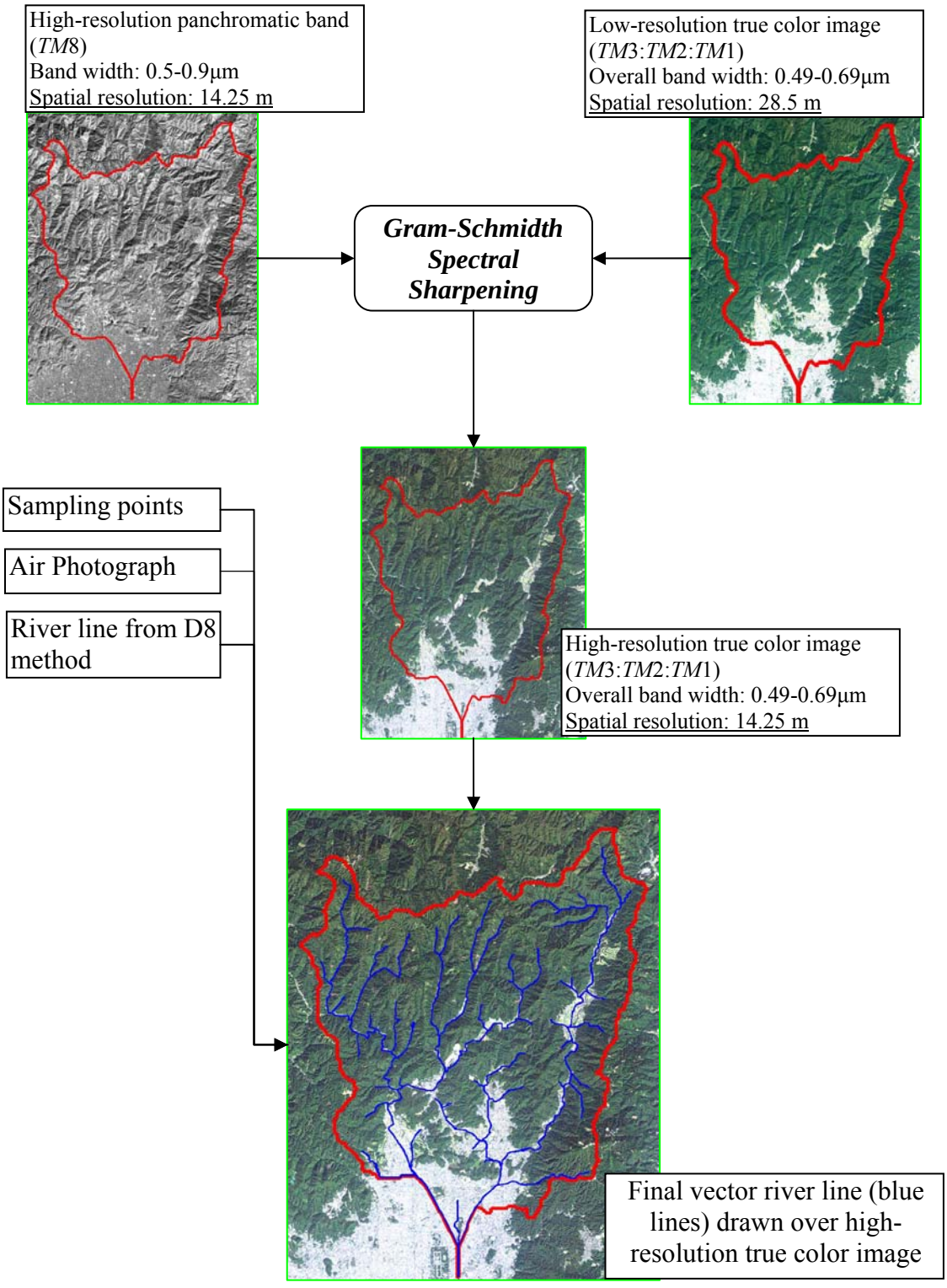


Figure 6.5 Flow chart showing the process to derive a river line over a transformed high-resolution Landsat ETM+ image of the study area

classes (< 3m, 3-10m, and >10m) we assumed them to be 1.5m, 6.5m and 15m respectively. Outlet heights of upper tank (S_{sf} and S_i) were then directly derived from the soil topographic index map as described in the earlier chapter (Eq. 5.17).

6.3 Results

6.3.1 Land cover classes

Figure 6.6 shows the decision tree used to derive the land cover class. In the decision tree, initial attempt was to separate the urban (non-vegetated) areas from vegetation areas. Since most of urban areas were relatively plain, so the vegetation in those areas was separated using the slope. The remaining areas were only hilly forest areas. Since images in three dates could have different shaded parts, so only the areas that were shaded in all images were classified as shaded vegetation. Using the information of *TM*5 and *TM*7 bands, three types of forest vegetations were separated, namely, deciduous, mixed, and evergreen in terms of descending magnitude of reflectance respectively. **Figure 6.7** shows land cover map along with their proportion by a pie diagram. More than 95 % of the areas were classified as vegetated areas, while non-vegetated urban areas accounted only 4.72 %. Vegetation class was further divided into five major sub-classes from the decision tree classifier. The sub-classes were namely, deciduous forest (18.07 %), evergreen forest (22.16 %) majority consisting of Japanese ceder (*Cryptomeria japonica*) and Japanese cypress (*Chamaecyparis obtusa* Endl.), mixed forest (M_{forest}, 40.32 %) consisting of Japanese ceder, Japanese cypress, and other deciduous species, shaded vegetations in hills (5.59 %), and vegetations near urban or plain areas (possibly shrubs, green tracts, golf courses, paddy fields, or other agriculture fields) (9.14 %).

Figure 6.8 shows the mean spectral pattern of vegetation classes. In all three images, differences in the reflectance of the vegetations could be observed with respect to the *TM*3, *TM*4, *TM*5, and *TM*7 bands. The reflectance pattern of *TM*3 and *TM*4 bands seems to be similar in all images, but the reflectance of *TM*5 and *TM*7 could be easily distinguished among them. The results showed the effectiveness of using seasonal images to classify the vegetation types inside the forest area. Besides, we also compared the land cover map and land use map, which was digitized from the paper map. **Table 6.2** shows the proportion of each land cover class (classified results) and land use classes. Forest area in the classified result increased slightly (~2%) compared with the land use map, whereas the water area

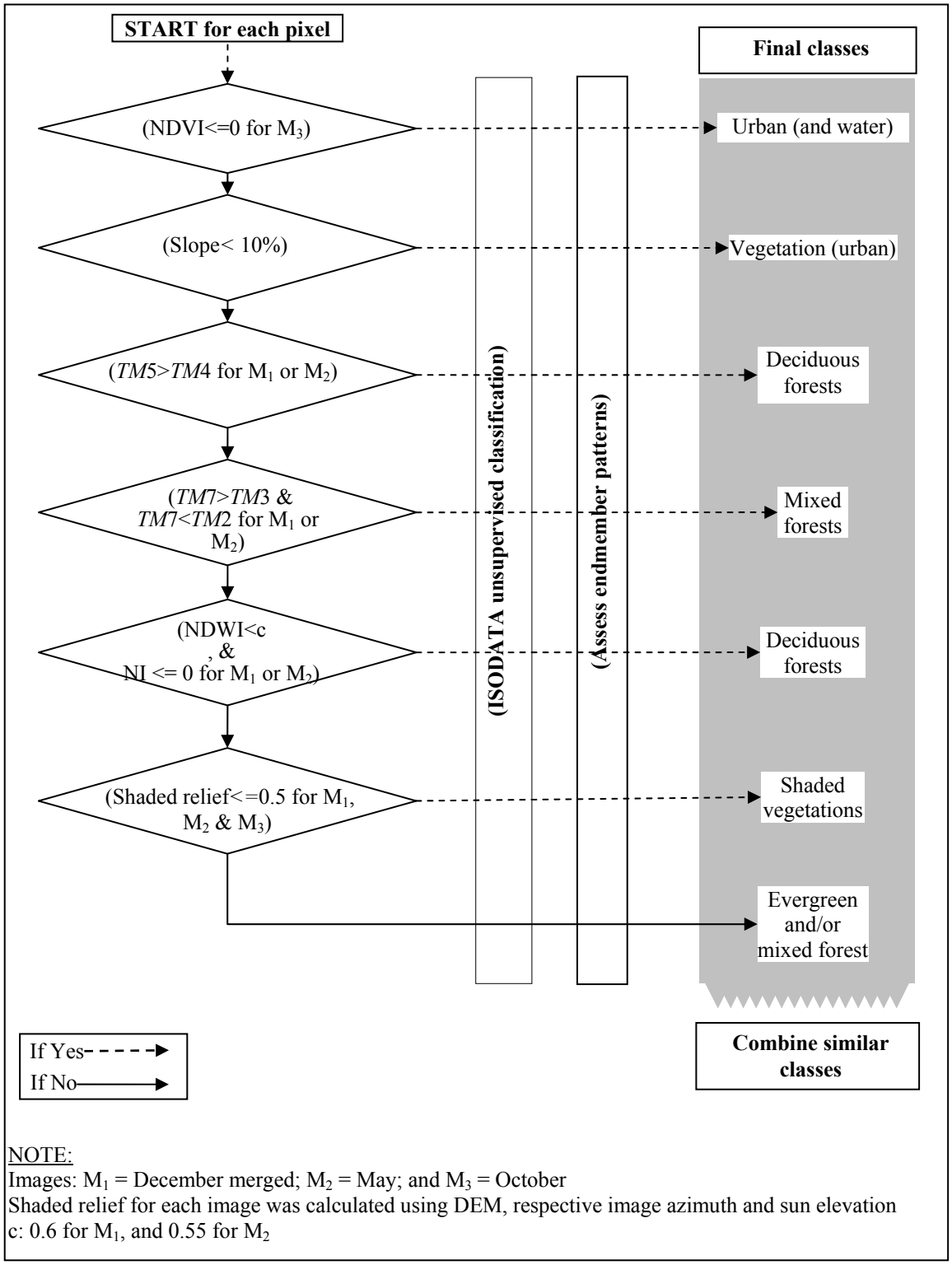


Figure 6.6 Binary decision tree used to classify the land cover of the study area

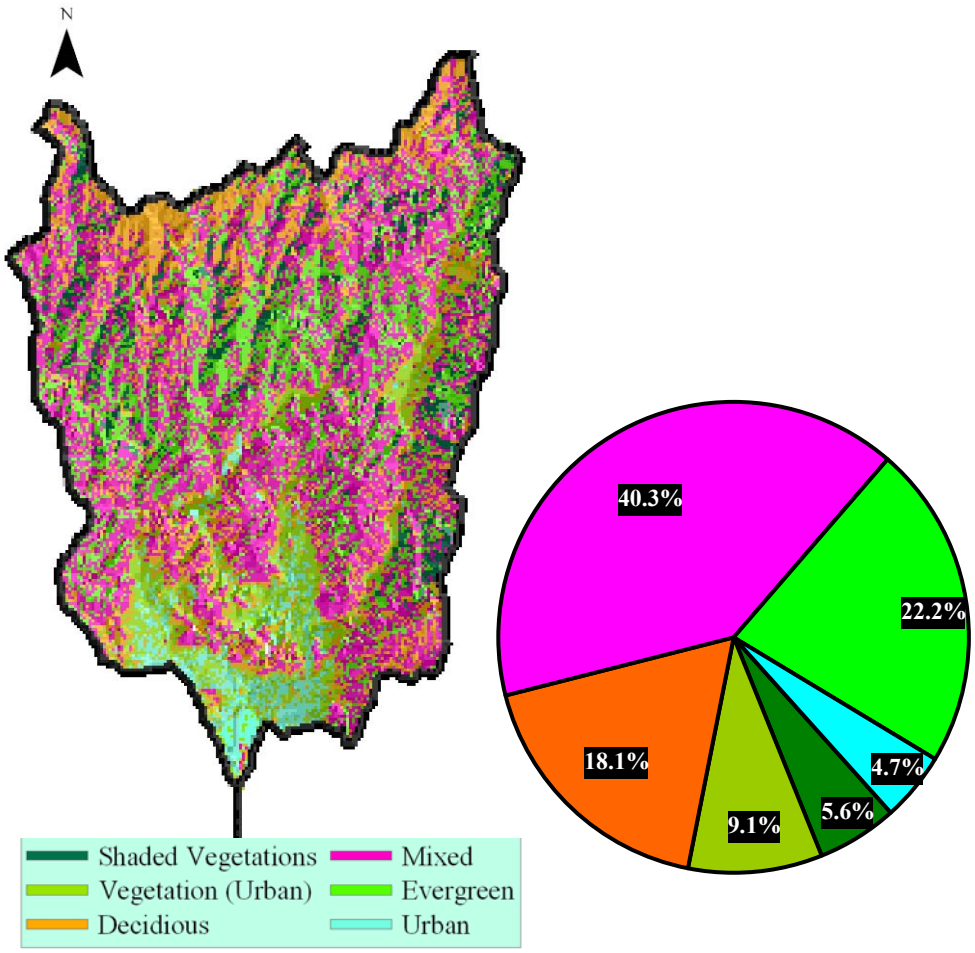


Figure 6.7 Land cover map showing the proportion of different categories

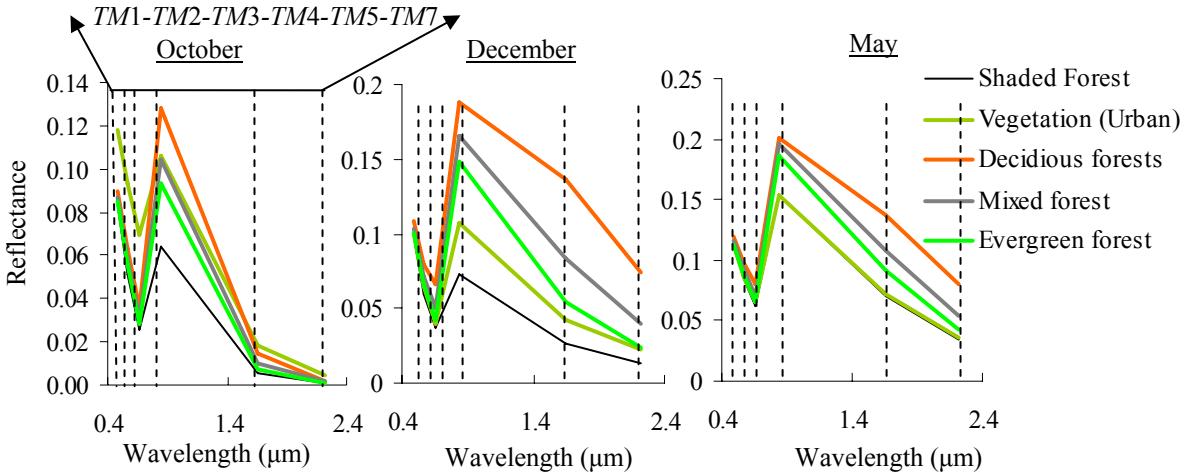


Figure 6.8 Spectral patterns of different vegetation types for different dates ETM+ images

Table 6.2 Land cover classification results and existing land use map classes cross comparison

Landuse (%) / Classified (%)		Water	Forests	Agriculture		Buildings	Residence	Green tracts	Bare land	**Total (%)**
				Paddy	Others					
Urban (+water)		0.14	0.45	0.34	0.02	0.18	2.85	0.1	0.01	**4.09**
Vegetation (urban)		0.48	2.25	1.62	0.05	0.18	4.37	0.19	0	**9.14**
Hilly Forest	Shaded	0	5.57	0.01	0	0	0.01	0	0	**5.59**
	Deciduous	0.15	15.63	0.69	0.04	0.1	1.73	0.32	0.04	**18.7**
	Mixed	0.08	38.64	0.32	0.03	0.07	1.03	0.13	0.02	**40.32**
	Evergreen	0	21.89	0.07	0.01	0.02	0.17	0	0	**22.16**
Total (%)		**0.85**	**84.43**	**3.05**	**0.15**	**0.55**	**10.16**	**0.74**	**0.07**	

was merged to urban class due to insignificant area. Small water bodies, such as small rivers, were undetected from Landsat images. Major mismatch between the land use map and classification results seemed to be due to the reflectance of vegetation. For instance, the non-vegetated areas in the land use map might actually contain significant vegetation. Such details were not be included in the land use map but easily detected from satellite image processing.

6.3.2 Leaf area index (*Lai*) distribution and canopy interception (*I*)

Figure 6.9 shows the map of *Lai* along with the histogram showing its distribution for the for each vegetation type. Among three forest vegetations, mean *Lai* was the highest for evergreen forest (average value: 2.9), which also had the range of 0 - 6. Mixed and deciduous forests had mean *Lai* of 1.86 and 1.76 and the range of 0 - 4 and 0 - 4.5 respectively. Higher *Lai* for evergreen vegetation could be due to their clumped nature, but

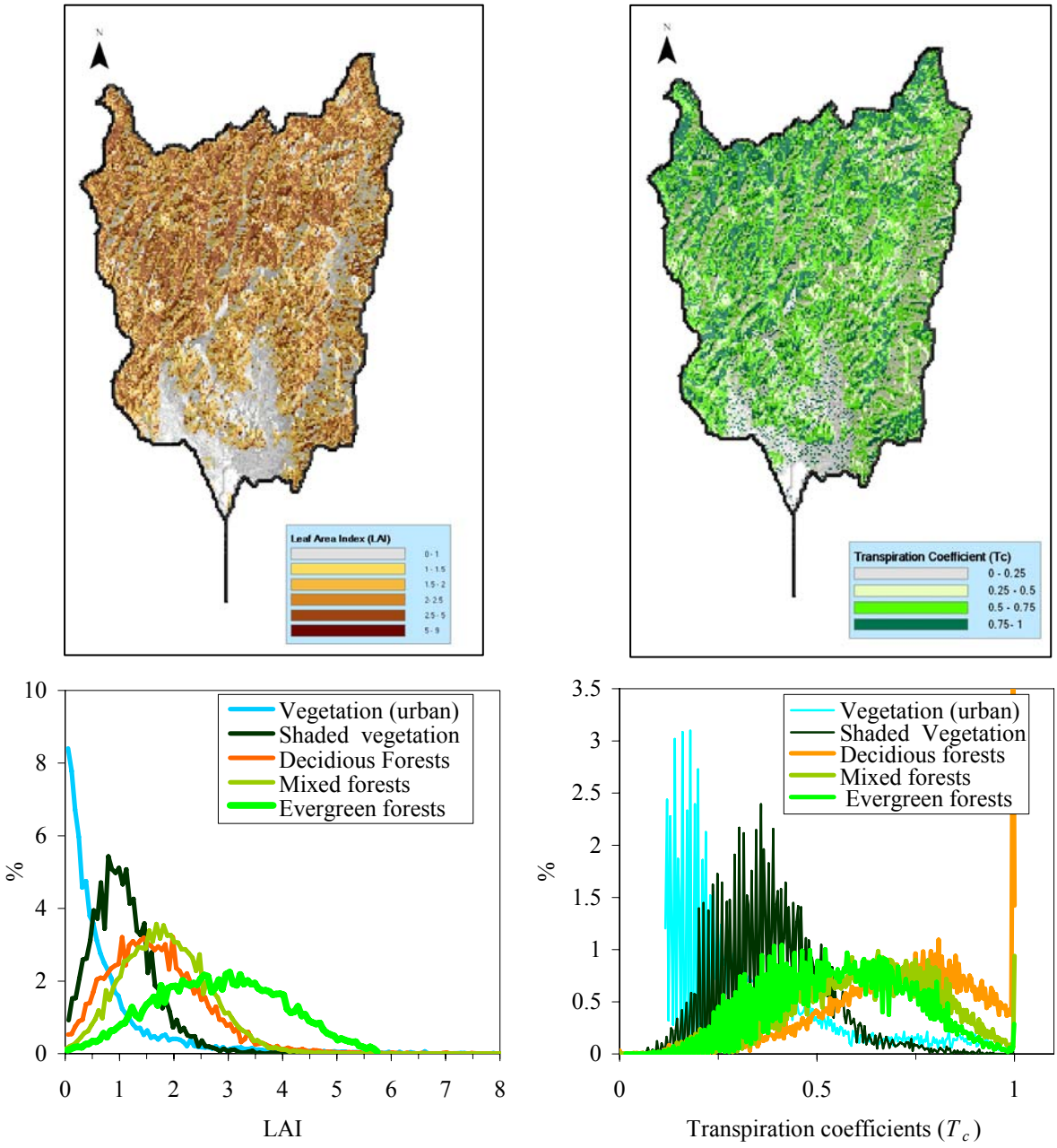


Figure 6.9 Map and distribution of *Lai* and T_c for different vegetation types

mixed forests contain both deciduous and evergreen species so their *Lai* was higher than deciduous. The maximum interception capacity of a grid could be quite higher (~ 1 mm) in the areas having dense forest cover. Due to dominant forest cover in the study area, significant amount of rainfall could be stored in terms of intercepted rainfall, especially, for smaller rainfall events.

6.3.3 Estimated evapotranspiration (*Et*) loss

Evapotranspiration is an important variable in modeling. The forest areas could significantly affect the evapotranspiration, so the value of T_c was quite important. Figure 6.9 shows the distribution of T_c for different vegetations, most of which seemed to be normally distributed. The mean T_c values for all three major vegetation types were greater than 0.55, which means the *Et* loss were generally more than half of Et_0. **Figure 6.10** shows Et_0 estimated by the model for the period of one year. It could be seen that there were wide variation in both daily Et_0 and mean temperature. To have a better outlook, the estimated *Et* was smoothened by the 15 days moving average, in which seasonal variation in the *Et* could be clearly observed. **Figure 6.11** shows the map of yearly *Et* loss from the watershed for 2003. *Et* ranged from 400 to nearly 1000 mm/year, in which most of the urban areas had the highest *Et* that was nearly equal to Et_0. It was mainly due to the preset criteria that in non-vegetated parts *Et* occur at potential rate due to free availability of the water. For Et_0, nearly 100 mm/year difference was observed between the maximum and minimum elevation grids. It could mean that consideration of effects of elevation on the evapotranspiration estimation was significant. During the summer months, Et_0 (~ 5 mm/d) was observed higher, but during the winter months it was less than 2 mm/day. In terms of vegetation types, most of the vegetations did not show much variation in *Et* loss, but they were only slightly lower than Et_0.

The estimated *Et* was also comparable with the one that was estimated in another forested watershed in Shiga prefecture, few kilometers northeast of the study area, in which daily estimated range was less 1 mm/d in winter but in summer period it was nearly 4mm/d (Kosugi and Katsuyama, 2007). In addition to that, in that study, average annual estimated *Et* was 750 mm/yr for 2003, which was quite close to the annual *Et* in our study area. If we compare the annual rainfall of 2003 (1814 mm) and *Et* loss, annual *Et* in some parts even reached nearly half of the actual rainfall in this study. Therefore, during the winter months,

Et loss might not have significant impact on the model estimation, but during the summer, *Et* loss could be significant to affect the model estimation.

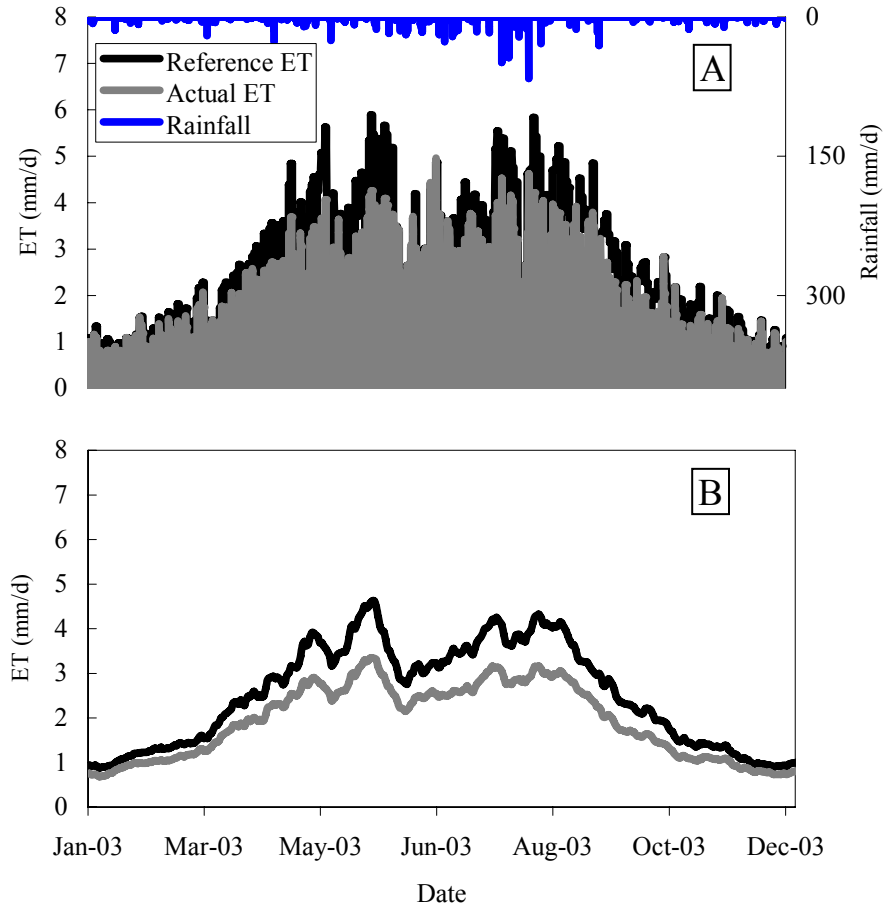


Figure 6.10 Daily *Et* estimation in the study area [Lower portion (B) represents the 15 days moving average applied to the upper portion(A)]

6.3.4 Distribution of slope and outlet height for surface runoff (S_{sf})

Slope (s) and S_{sf} (or S_i) were important parameters of upper tank. **Figure 6.12** shows the map and distribution of s and S_{sf}. It could be seen that majority of the hilly areas (~ 80%) were in the upstream parts that usually had elevation range of 300-900 m (Fig 3.7(a)). Majority of slope (~ 60% area) were greater than 0.4, which indicates that the influence of topography must be quite higher on the surface and sub-surface runoff. If we consider the distribution of S_{sf}, in the study area, it could be seen that nearly 80 % of the areas had higher S_{sf} (> 10 mm), which was similar to the proportion of normal grids (84%). Therefore, overland runoff processes in the normal grid areas could be quite limited in comparison with the subsurface runoff or percolation to lower tank. Therefore, only

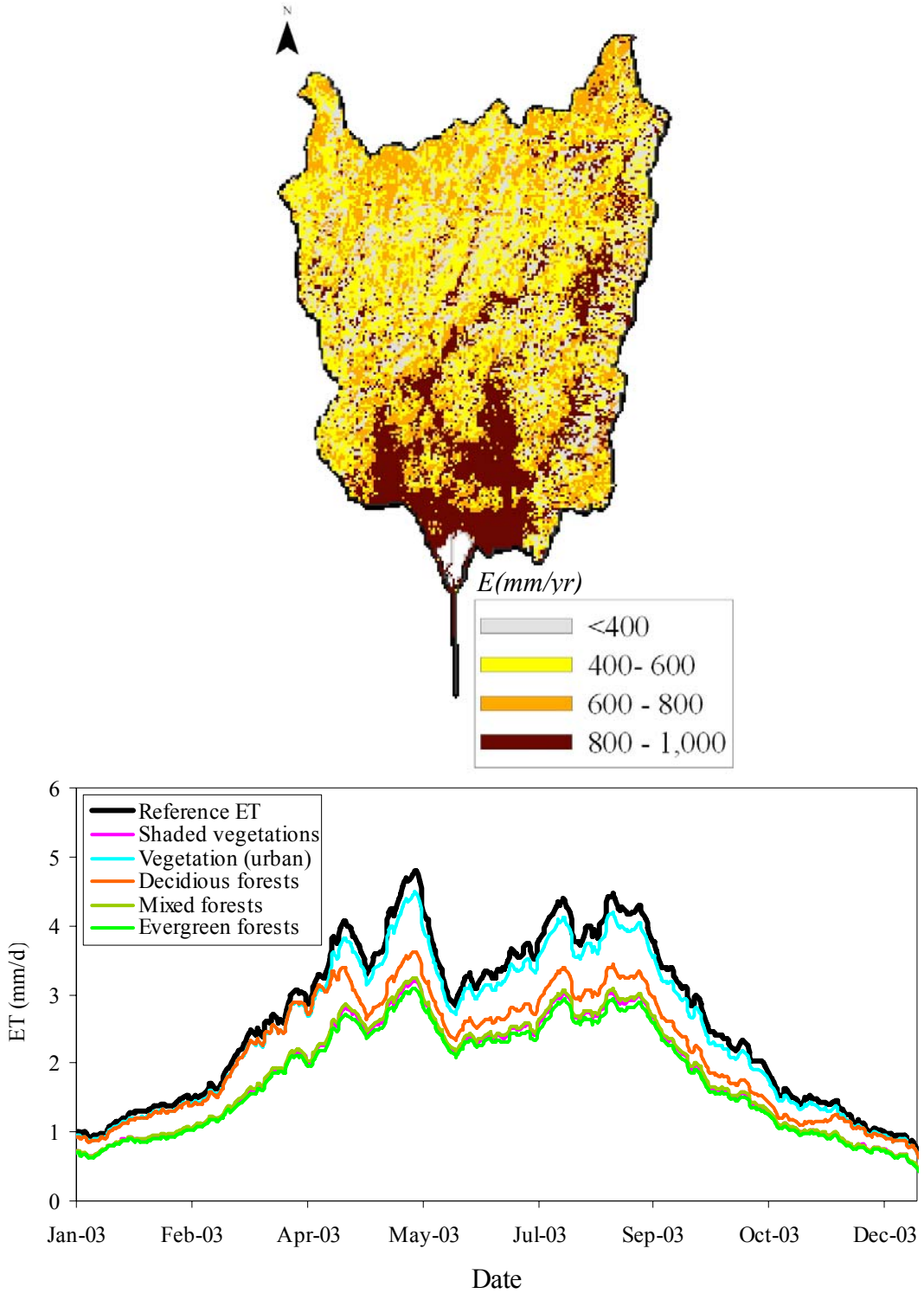


Figure 6.11 Annual *Et* map of the study area (upper) and *Et* loss for the year 2003 by different vegetation (lower, smoothened by applying 15 days moving average)

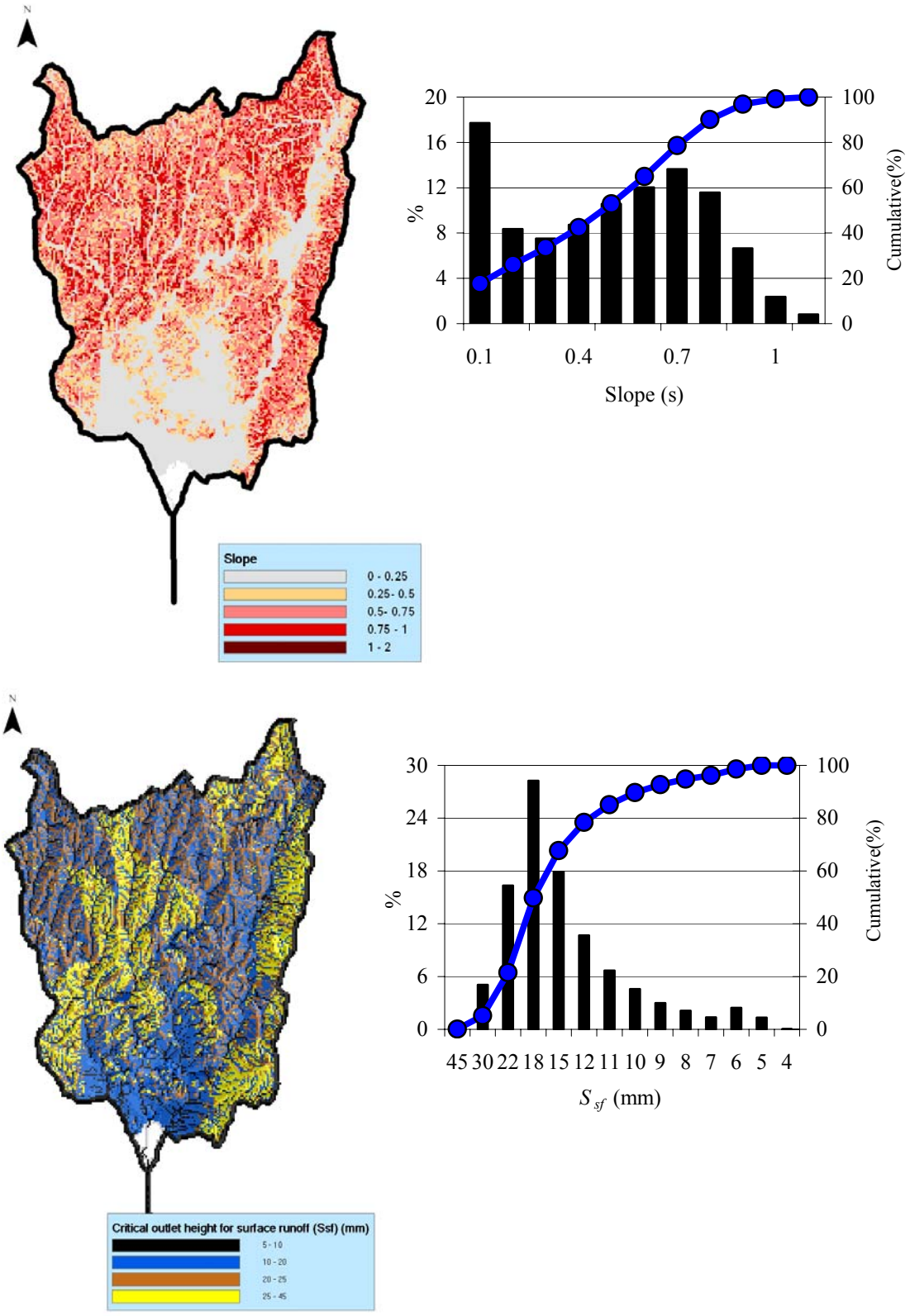


Figure 6.12 Distribution of slope (s) and outlet height for surface runoff (S_{sf}) in the study area

channel grid or nearby normal grids could contribute the overland runoff especially during drier or lower intensity rainfall events.

6.4 Summary

Remote sensing and GIS have enhanced our understanding of spatial phenomenon to a vast extent. Both have emerged as potential techniques that could significantly improve the effectiveness of distributed models to estimate runoff and water quality. In this chapter, remote sensing and GIS techniques were used to derive watershed attributes that were indispensable both during the design of the model and to assign the values of the distributed parameters. Important aspects of this chapter are summarized by the following points:

1) Three seasonal Landsat ETM+ images were processed to derive the land cover map of the study area including major forest types (deciduous, mixed, and evergreen);

2) With the help of land cover map and by the use of reflectance of two thematic bands (*TM*3 and *TM*4) of Landsat ETM+ green leaf area index (*Lai*) was determined, which was required for the estimation of canopy water storage by the model;

3) Transpiration coefficient of green vegetation was determined with the help of *TM*3 and *TM*4 bands. It was an essential parameter to estimate the evapotranspiration loss;

4) Evapotranspiration (*Et*) was estimated by FAO Penman-Monteith method, by which in an average 5 mm/d during summer period and less than 2mm/d of *Et* loss during winter was estimated;

5) Important topographic variables, such as, slope, drainage direction, upslope contributing area, and outlet height for surface runoff were determined mainly by the processing of DEM using GIS techniques;

CHAPTER VII

APPLICATION OF THE DISTRIBUTED WATER QUANTITY AND QUALITY MODEL IN THE STUDY AREA

7.1 Introduction

Model assessment shows the effectiveness of the model to predict the changes at different conditions at a given place. Central idea in the model assessment is to examine how values of a given set of parameters would give reliable outputs at varied rainfall-runoff conditions. Flow rates or concentration at certain points of the watershed generally constitutes major outputs of the model. With few exceptions, only option to assess the model performance is to compare between observed data and estimated outputs during the calibration or validation of the model. However, comparison between observed and estimated data may not be always effective to ensure the applicability of distributed models. In case of distributed models, it is also likely that changing size of the grid could affect the values of model parameters thereby showing variation in the estimated results (Vieux and Needham, 1993; Kuo *et al*., 1999; Vazquez *et al*. 2002). Therefore, it is quite essential to assess the effects of grid size on the model performance because changing grid size is also accompanied by the changes in the watershed attributes.

This chapter attempt to evaluate the distributed water quantity and quality model, designed in Chapter V, firstly through the comparison between the estimated results and observed data. Secondly, effects of changing grid size on the model estimation as well as the values of essential parameters will be assessed. Finally, the variation in estimated results at basic grid size will be examined at different possible future scenario.

7.2 Methodology

The model at basic grid size (~ 50m) was simulated by changing its parameters to fit with the observed data on different spatial and temporal conditions. Initial attempt was to estimate the water quantity, such as river flow rate, because its estimation would directly affect the estimation of the water quality indices (WQIs). Estimation of flow rates was

followed by the estimation of the selected WQIs through the adjustment of essential parameters. Since the main purpose was to show the potential of distributed model to estimate the river water quality, we selected two common WQIs (COD_{Mn} and TP) in our study. If the estimated results were found satisfactory, the values of parameters were assumed standard values for the 50m grid size. The estimation of the basic grid size was further examined on the basis of possible changes in the watershed attributes (such as rainfall and land use patterns) in the future.

The standard values of parameters for 50m grid size were used to simulate the model for other nine grid sizes. On the basis of the estimated differences, model was re-calibrated by adjusting some critical parameters for nine grid sizes in order to assess the effects of grid size on the values of the parameters. Rather than seeking for precise re-calibration, effort was put on to understand the pattern of values of selected parameter at different grid sizes.

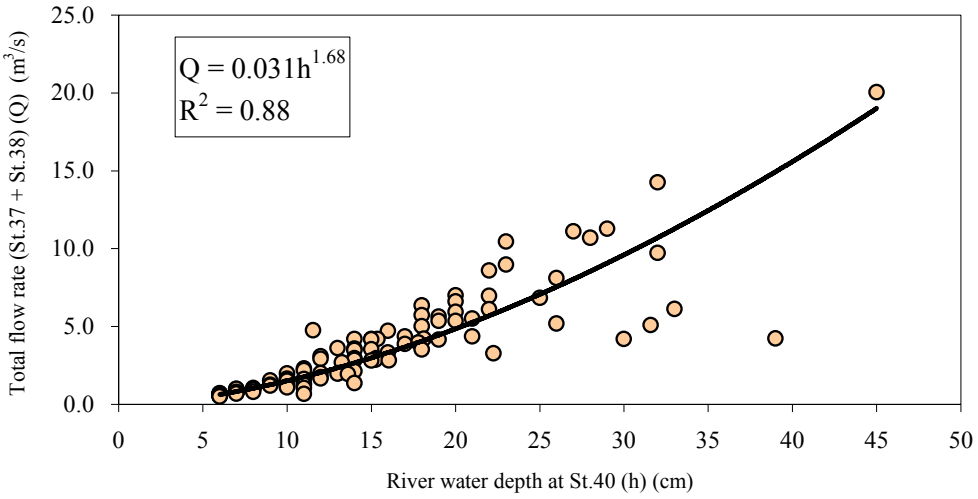


Figure 7.1 Relationship between the observed flow rates (St. 37 + St.38) and river water depth (St.40).

7.2.1 Model assessment with the help of observed data

At first model was applied to estimate the water and pollutant discharged at all 39 stations, which was mainly focused to see the model response at base flow conditions (no direct rainfall condition). Simultaneous survey data were used to assess the model estimation at 39 stations. Simultaneous survey data represented the condition at the time of sampling, so continuous simulation data was used to assess the model performance at different runoff conditions. We firstly used daily river water depth, observed at the outlet station (St. 40), to access the model performance. The river water depth was converted to

flow rates by performing regression between the observed flow rate (St.37 + St.38) (Q) and river water level data (h) (cm) as shown in **Fig. 7.1**. The relationship between Q and h was shown by the power equation in which R^2 was 0.88. Besides, observed flow rate obtained from different temporal surveys (alternate days, every 10 days, storm event) were also used to assess the model performance, especially focusing on the observed runoff patterns.

Several evaluation criteria were available to assess the model output based on the Least Square Difference as explained in Chapter II. In this study, we focused mainly on two commonly used criteria, namely, root mean square error (*RMSE*) and relative mean square error (*RE*) for the assessment of estimation by the model, which were shown by **Eq. 7.1** and **7.2**.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - E_i)^2} \qquad \textbf{7.1}$$

$$RE = \frac{\sqrt{\frac{1}{N}\sum_{i}^{N}(O_i - E_i)^2}}{\frac{1}{N}\sum_{i}^{N}O_i} \qquad \textbf{7.2}$$

Where, O_i and E_i were the observed and estimated values at i^{th} hour. *RMSE* had same unit as of the interest variable, but *RE* was unit less. All above evaluation criteria were named as "normal" method because they considered only one estimated point relative with one observed point. However, majority of observed data were taken at different time scale, which were usually greater than output time of the estimated result (*i.e.*, one hour). In such condition, even though the magnitude of peak points of observed and estimated were same, it was likely that they differ by some time (Sugawara, 1995). Therefore, minimum squared difference in either of *i*-1, *i,* or *i*+1 hour estimated data was additionally used in above equations (7.1 ~ 7.3) to evaluate the model estimation:

$$(O_i - E_k)^2 = Minimum\left\{(O_i - E_{i-1})^2, (O_i - E_i)^2, (O_i - E_{i+1})^2\right\} \qquad \textbf{7.3}$$

Where E_k is the estimated values in either of *i*-1, *i,* or *i*+1 hour showing minimum squared difference with the observed value (O_i). The evaluation criteria (*RMSE* & *RE*) were calculated by using **Eq. 7.3,** and termed as "modified" method as suggested by Sugawara (1995).

7.2.2 Criteria for grid resizing

According to second objective of this chapter, spatial data of different grid sizes were required. In distributed modeling, it hardly coincides that all data were available at the desired scale required by the model. Because of the unavailability of original data at the desired scale, it was mandatory to resize the base grid of 50m to other grid sizes. For this study, nine more grid sizes (100, 150, 200, 250, 300, 350, 400, 450, 500 m) were derived by multiplying the sides of 50m (46.21m × 56.91m) grid size by 1,2,3,…,10 respectively.

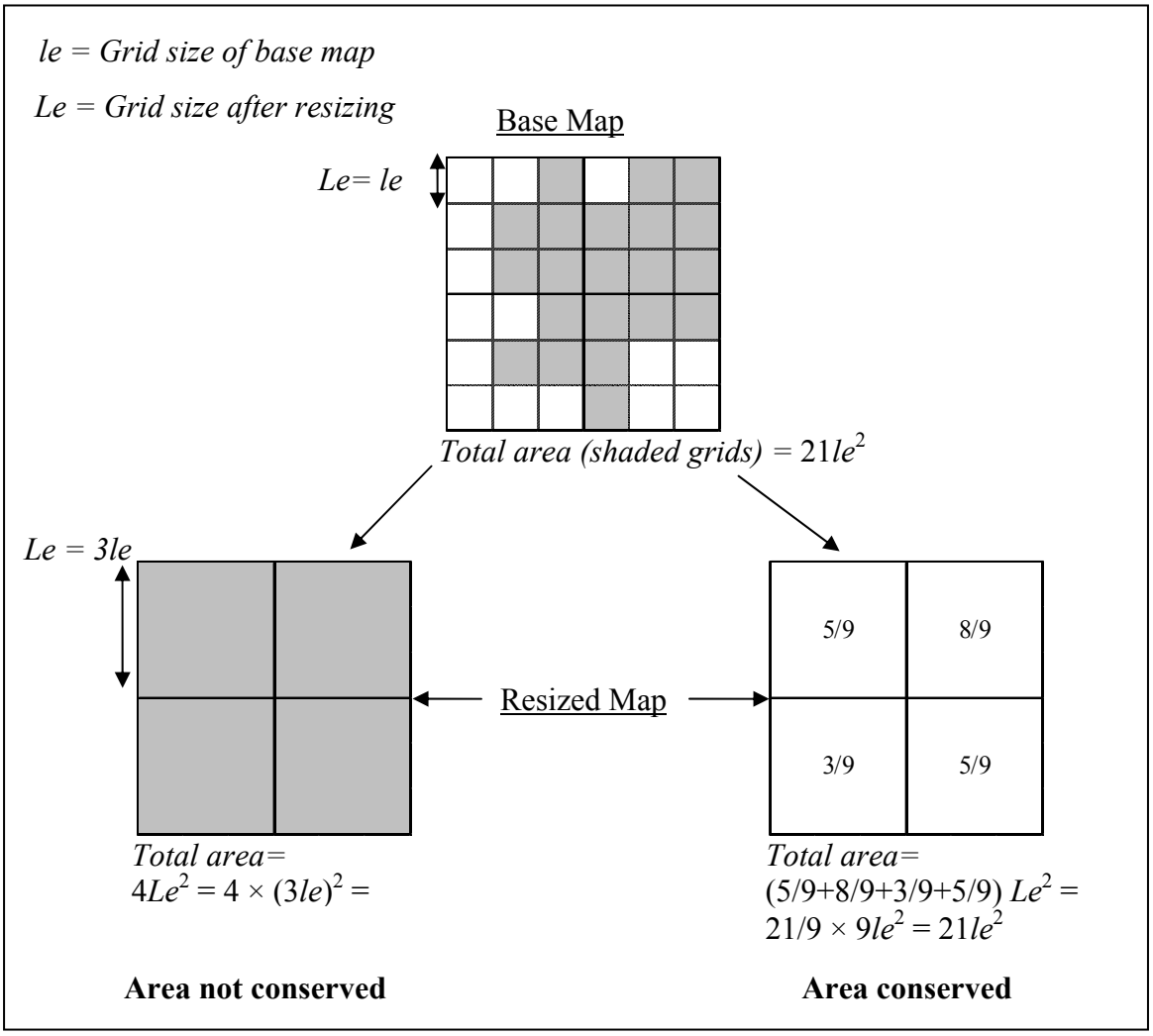


Figure 7.2 Difference between the grid resizing techniques by conserving and without conserving the watershed area at the boundary.

During resizing of basic grid size, appropriate techniques were adopted, depending on the types of spatial data (categorical or continuous), to minimize the effects of grid resizing on

watershed attributes. At first, it was necessary to ensure that no significant information was modified during the resizing of the grid. One of the serious effects of grid resizing could occur on the area of whole watershed, if the boarder cells were not adjusted for the area inside true boundary (or base map) (Bruneau *et al*., 1995). To adjust such effects, we used a GIS technique to correct the area of border cells as shown in the **Fig 7.2** (Kuo *et al*., 1999). If shaded grids ($21le^2$) represented the watershed area inside the boundary, then white grids were the area outside the boundary of a base map. If corrections were not done, the total area of resized grid (Le) would be $36le^2$, which was greater by $15le^2$ than in the base map. If corrections were done, only shaded fraction would be considered during calculations even though the resized grid had larger area. In that way, total watershed area would be always conserved even after the resizing of the grids. Therefore, area of resized grid size could be multiplied with a correction factor (A_c) to conserve the area of basin:

$$A_c = \frac{A_f}{A_T} \qquad \textbf{7.4}$$

Where, A_f was the actual area in base grid size contributing to resized grid, and A_T was the area of the resized grid irrespective of the actual area in the base map. Population was added in the resized grid so that no change in the number of population occurred in any grid sizes. Resizing of grids generally resulted population from normal grids merged into the sewer category, so the population loadings were accounted irrespective of the grid types in the resized cases.

Categorical data such as land cover and geology (or soil profile depth) were resized according to the majority (>50%) rule. When majority rule was not possible, the category for the resulting grid was based on the nearest neighborhood. The continuous variables, such as DEM, *Lai*, and T_c were resized by using average of corresponding grid cells. The resized DEM maps were further processed to derive the topographic variables, such as slope, drainage direction, upslope contributing area, and soil topographic index as described in the earlier Chapter VI (Fig. 6.3).

7.2.3 Model assessment at different grid sizes

At first, parameter values obtained in the 50m grid were applied to the model for all grid sizes data. In addition to the difference between the observed data and estimated

results, difference between the estimated values obtained from 50m grid size and other grid sizes were considered. For example, *RMSE* was calculated as shown in **Eq. 7.5**. Where,

$$RMSE = \left(\sum_{i=1}^{N} (Eb_i - Eo_i)^2 \Big/ N \right)^{0.5} \qquad \textbf{7.5}$$

Eb was the estimated output at 50m grid size, and *Eo* was the estimated output at other grid sizes. Since the estimated values by the model at 50m grid size was supposed to show satisfactory, so the comparison would provide an additional measure to assess the differences.

Based on the estimation differences shown by the model at different grid sizes at a given value of each parameter, model was separately calibrated for all grid sizes. Main purpose of the calibration was to assess the effect of grid size on some essential parameters. Selection of parameters for calibration was done after assessing the effects of each parameter on the model estimation. It was done by changing the values of a particular parameter while values of other parameters were kept constant. Using selected parameters, the calibration of the model for each grid size was performed until the difference between the estimated values of 50 m grid size and remaining grid sizes became minimum, which was done by repeated simulation starting from wider range of parameter values.

Besides descriptive error criteria, visual inspection of the hydrographs or pollutographs was also emphasized in the study, especially focusing on the magnitude of peak, duration of peak event, time required for reaching the peak point, and time required for the recession of the peaks.

7.2.4 Evaluation of model estimation at different scenario

Distributed models are flexible in comparison with the lumped models in that effects of specific changes in some parts of the watershed could be independently incorporated in the modeling framework. The distributed models are more appropriate to assess the likely changes in the future, such as land use differences, change in rainfall patterns and so on. Among the changes land use changes are most dynamic, such as by urbanization, deforestation, change in cultivation practices, introduction of new sewer systems, population migration *etc*. Firstly, we mainly focused on the changes in the rainfall possibly due to the changes in the global climate in future. Secondly, we assessed the effect of land

cover changes, such as by increased urbanization. The major aim of such simulation was to evaluate how differences could be observed in the estimated results by major changes in distributed points.

Global warming has been seen as a new issue that is often concerned for unprecedented changes in the hydrological cycle. In case of Kyoto, in 125 years (1881-2006) temperature showed continued rise, with net increase of nearly 4^0C in the yearly average temperature. It could be possible that in future watershed might face drought or extreme flooding. Although remarkable difference was not seen on the rainfall pattern during past 125 years in the study area, the minimum (880 mm/yr) and maximum (2180 mm/yr) of annual rainfall observed during the periods were considered as extreme events. The data were only available for monthly basis, so we assumed the rainfall distribution pattern on 2003 was representative in the study area.

$$R_i = \frac{(r_{main-hourly})_i}{r_m} R_m \qquad \{i = 1,2,3.....24;\ m = 1,2,3,....12 \qquad \textbf{7.6}$$

Then **Eq. 7.6** was used to transform the observed data of 2003 to corresponding data for extreme events. Where, R_i was the transformed rainfall (mm/hr) for given extreme event data. $r_{main\text{-}hourly}$ was the rainfall observed at i^{th} hour and r_m was the total rainfall for m^{th} month of the year 2003, and R_m was monthly rainfall for either of minimum (800) or maximum (2200) annual rainfall observed during 1881-2006 at the main station. The main aim for this assessment was to examine that in the event of extreme condition how the river runoff situation and water quality would become.

To assess the effect of land cover changes, slope and elevation were used to select the areas that might possibly undergo urbanization. Three conditions were devised.

Con1: Areas less than 300m and less than 30% slope were assumed urban areas;

Con2: Areas less than 500m and less than 40% slope were assumed urban areas;

Con3:

a) Areas of urban and vegetated urban classes were kept same;

b) Rest of the areas less than 300m and with slope less than 20% were transformed to vegetated urban class;

c) Rest of the areas less than 500m and with slope less than 40% were transformed to urban class;

In above conditions, the remaining areas were kept unchanged. Although the conditions were hypothetical, the estimated values could provide the extent of variation on the river water quantity and quality by the likely changes in future.

Table 7.1 Values of the parameters of hydrological sub-model used in this study

Class combination*	FH	FM	FL	V	U
f_1(1/hr)	0.225			0.338	0.450
f_r(1/hr)	0.25				
f_0(1/hr)	0.00486			0.000486	0.000486
g_1 (1/hr)	0.00012	0.00008	0.00004	0.00004	0.00004
S_g	180				

* 1st and 2nd letters (e.g., FH) are land cover and permeability type, where, F: forests; U: urban; V: vegetation (urban); H: high; M: medium; and L: low

Table 7.2 Values of the parameters of water quality sub-model used in this study

WQIs	Accumulation rate (k_0) (kg/km^2/d)			Person (g/d)	Maximum buildup possible (w_{max}) (kg/km^2)	Sweeping coefficient (a)	Washoff exponent (c)
	Forests	Urban	Vegetation (urban)				
COD_{Mn}	20	40	60	7.7	$40 \times k_0$	0.008	2.45
TP	0.354	2.080	0.658	0.75	$100 \times k_0$	0.001	1.95

7.3 Results

7.3.1 Values of parameters

Table 7.1 and **Table 7.2** shows the values of the parameters of hydrological sub-model and water quality sub-model respectively, which was initially used to simulate at 50m grid size model. Although recommended values were used, they were modified to fit the estimated results with the available observed data.

7.3.2 Model estimation

At 39 stations:

Figure 7.3a showed the estimated flow rate in the 39 sampling stations. Majority of observed and estimated results were in quite close range. From the upstream to downstream, the observed and estimated flow rate showed similar pattern. Similarly, comparison of observed and estimated data for all stations also showed proportionate relationship having higher correlation coefficient (r = 0.9). The estimated pattern might indicate that the model response at the outlets during base flow condition could be satisfactory. Since the simultaneous surveys did not represent continuous data, it could be

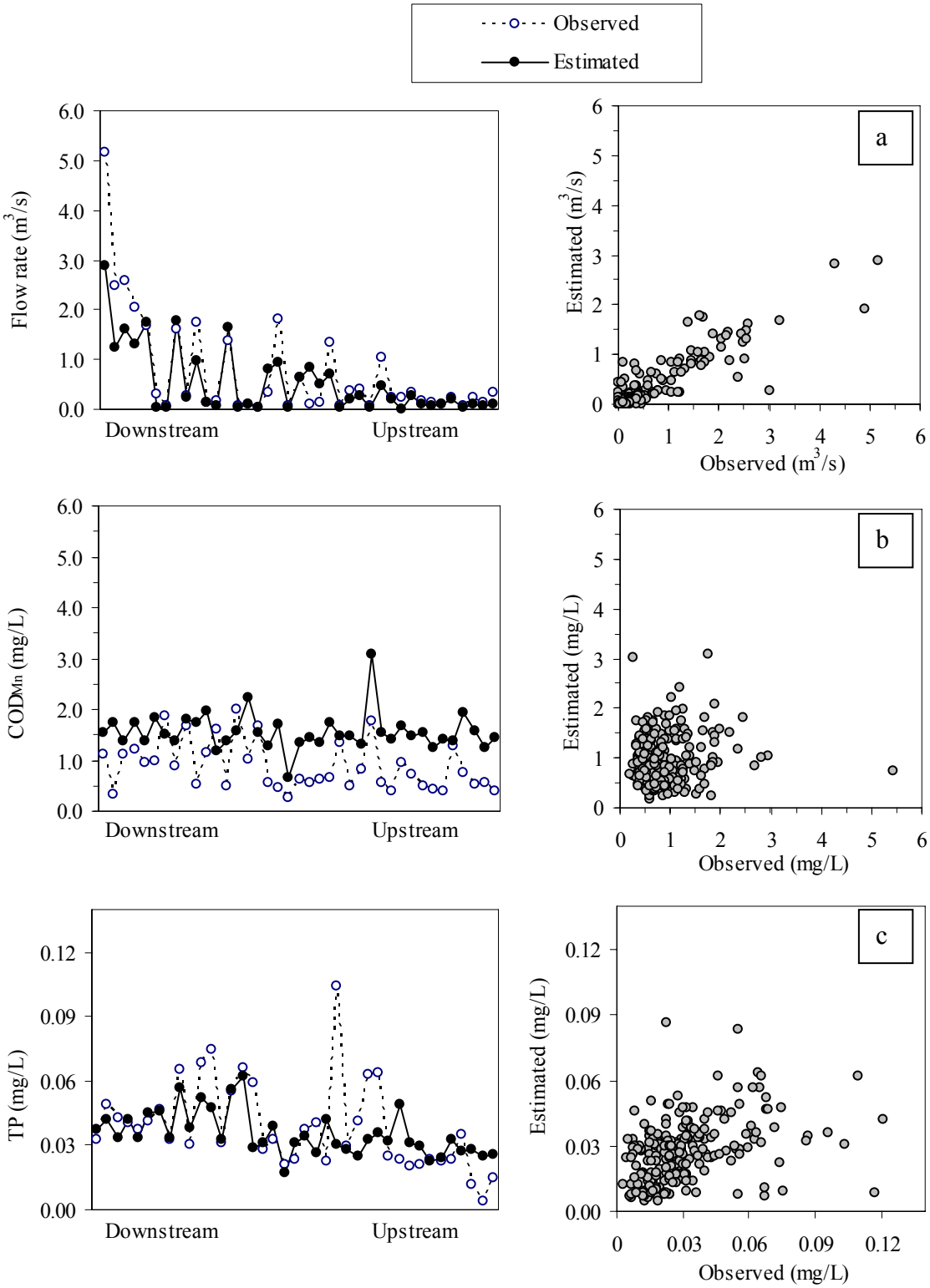


Figure 7.3 Comparison between observed and estimated pattern of flow rate, COD_{Mn} and TP at 39 sampling stations. (Right part shows the data of 2001/10/24 survey;and left portion shows the comparison for all surveys data)

considered as preliminary assessment of the model, especially, indicating that model might estimate within reasonable range at distributed points.

Estimated flow rates at different temporal conditions:

Figure 7.4 shows the comparison between the daily observed and estimated flow rates at the St.40 for the period of 2002, September to 2003, December and **Table 7.3** shows the values of selected evaluation criteria. Model estimation well followed the observed pattern, except some estimated peaks showing slightly higher values than the observed data. Particularly the short duration steep peaks were well simulated by the model. The *RMSE* was only 3.8 m^3/s (1.6 m^3/s by modified method) which was quite lower as compared with maximum observed flow rate that was more than 43 m^3/s (~ 74 cm river water depth). Similarly, the *RE* was 0.81 indicating that error was less than the average of observed flow rates. Although the observed daily flow rate was derived from the regression, the similarity of the estimated flow rate indicated that the model could be effective for the continuous simulation. Due to the lack of direct measurement of flow rates at St.40, the evaluation could not be considered as the final assessment. For further assessment, we utilized the flow rate observed during temporal surveys (St.37 and St.38).

Figure 7.5 and **Figure 7.6** shows the model estimation at St.37 and St.38 respectively, in which alternate days and storm events surveys are shown separately. Estimated results well followed the observed data for both stations. The overall estimation of model also showed quite satisfactory range as shown by lower *RMSE* (1.95 m^3/s for St.37 and 1.7 m^3/s for St.38), in which modified method gave still lower value in almost cases. As mentioned earlier, the data represented different types of surveys so it was necessary to evaluate the estimated results in terms of survey type, namely, alternate days, 10 days, and storm events. In alternate day surveys, majority of the observed data resemble the base flow conditions, except few periods that received rainfall intensity of 26 mm/d. Although the estimated patterns were similar with the observed data, *RE* was highest among all indicating the influence of majority of observed base flow periods than estimated peaks. The observed and estimated patterns for the 10 days observation however showed similar *RMSE* as compared with the all data. In case of storm event surveys, *RMSE* was observed quite higher in spite of very similar prediction with the observed data. In storm events, the magnitude of observed and estimated data was quite big, which could cause higher *RMSE*.

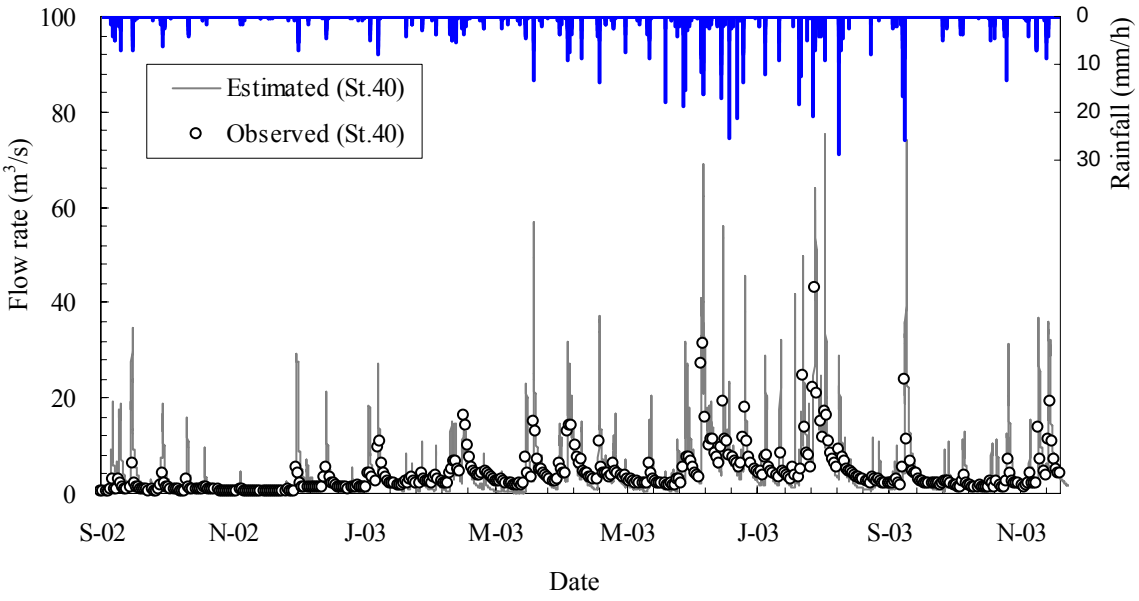


Figure 7.4 Comparison between observed and estimated flow rates observed at St.40 for the whole simulation period.

Table 7.3 Performance of the model for the estimation of flow rates shown by the values of selected evaluation criteria

Survey/data types	Evaluation Criteria	Stations							
		St.40#		St.37		St.38		St.(37+38)	
		Normal	Modified	Normal	Modified	Normal	Modified	Normal	Modified
All data	*RMSE*	3.28	1.61	1.95	1.72	1.70	1.30	1.82	1.51
	RE	0.81	0.40	0.77	0.68	0.72	0.55	0.74	0.61
Alternate days	*RMSE*	3.43	0.89	1.59	1.40	0.94	0.59	1.27	0.99
	RE	3.38	0.88	2.79	2.45	1.71	1.07	2.25	1.76
10 days	*RMSE*	2.67	1.25	1.23	1.16	1.17	1.13	1.20	1.15
	RE	0.70	0.33	0.66	0.62	0.60	0.58	0.63	0.60
Storm events	*RMSE*	4.32	1.56	2.47	2.17	2.31	1.72	2.39	1.95
	RE	0.92	0.33	0.57	0.50	0.59	0.44	0.58	0.47

Evaluation for St.40 were calculated with the help of flow rate (Q) derived from the river water depth (h)

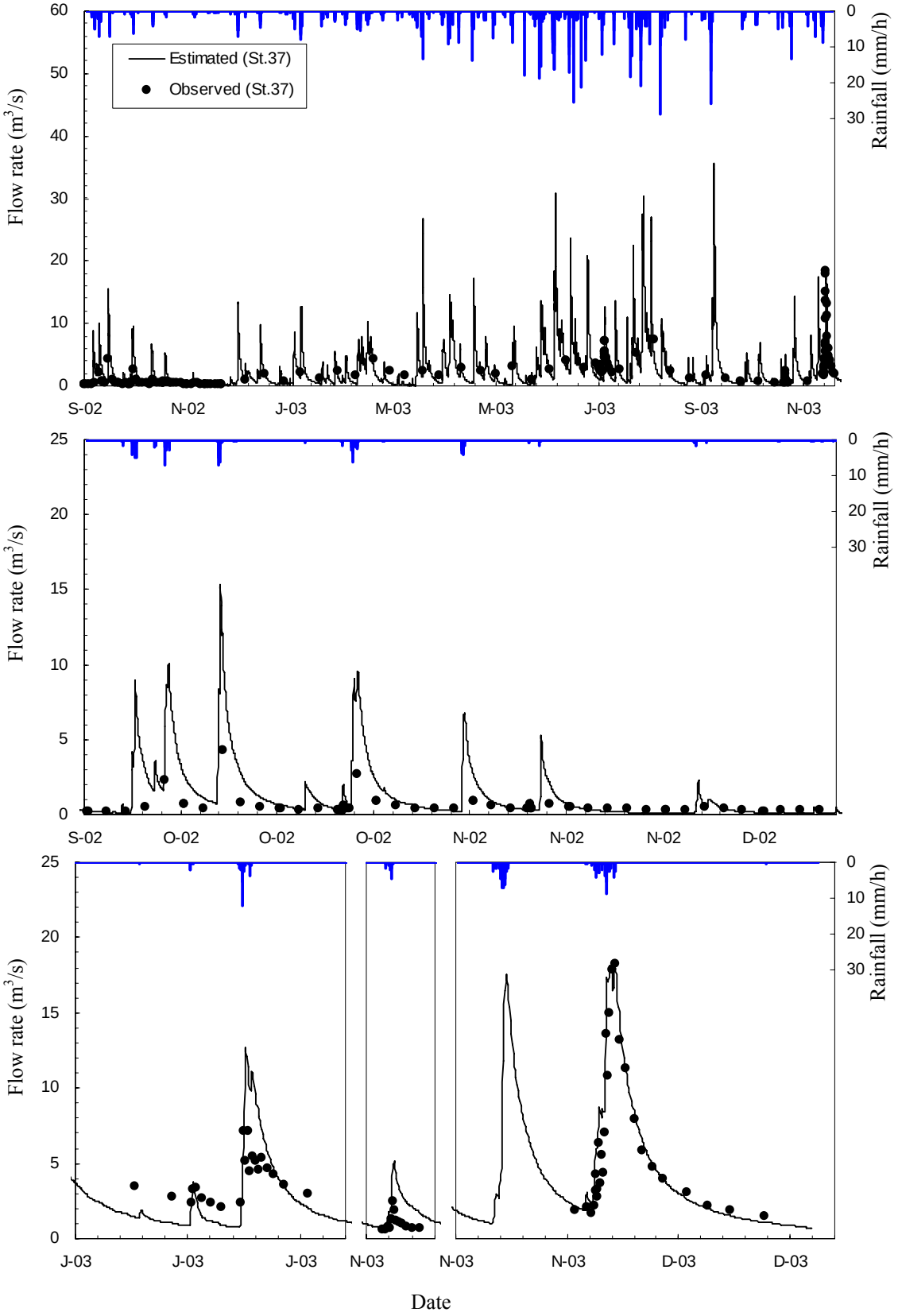


Figure 7.5 Estimated and observed flow rates at St. 37

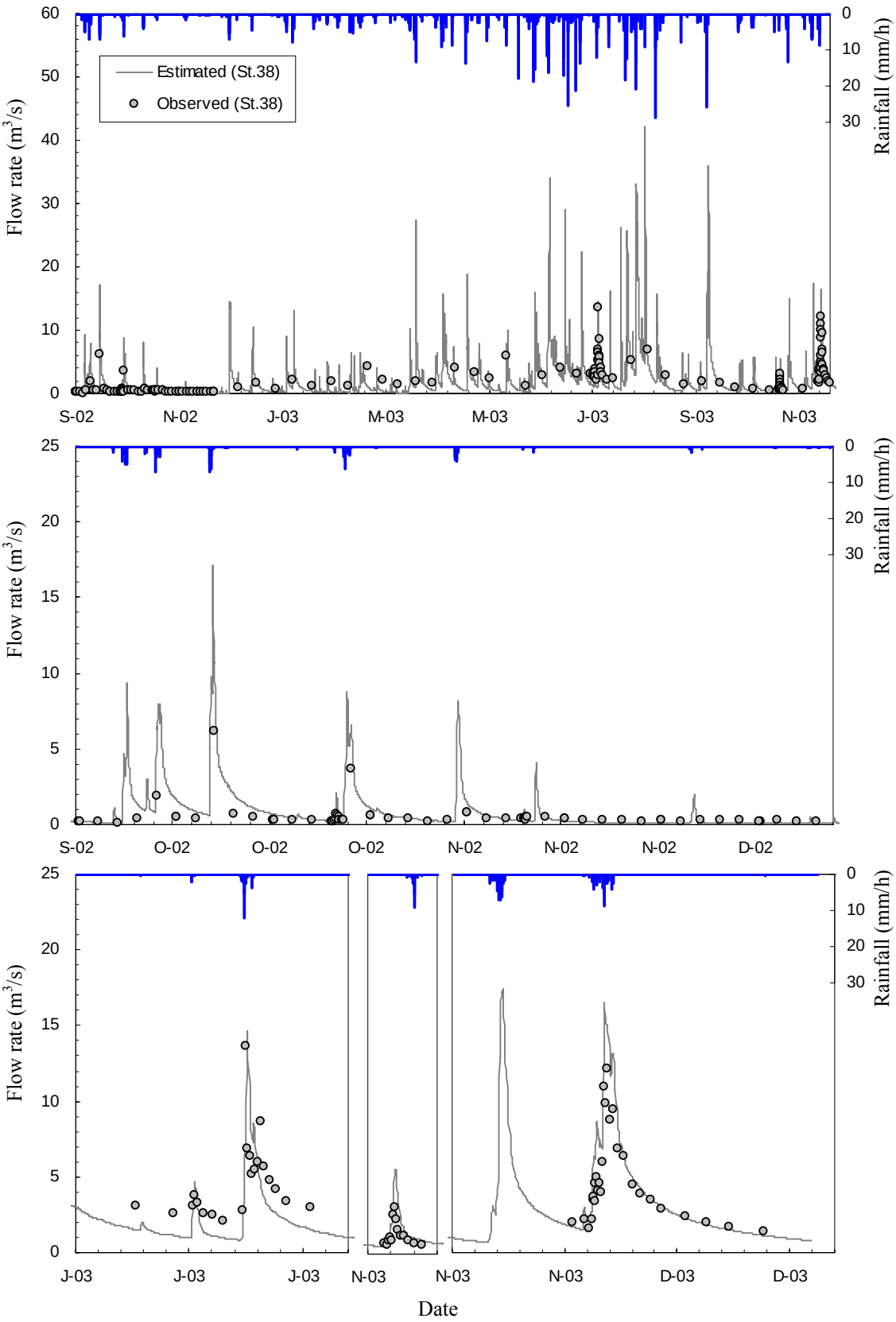


Figure 7.6 Estimated and observed flow rates at St. 38

Table 7.4 Performance of the model for the estimation of COD_{Mn} concentration shown by the values of selected evaluation criteria

Survey/data types	Evaluation Criteria	Stations					
		St.37		St.38		St.(37+38)	
		Normal	Modified	Normal	Modified	Normal	Modified
All data	*RMSE*	1.21	1.10	1.41	1.27	1.31	1.18
	RE	0.67	0.61	0.82	0.73	0.74	0.67
Alternate days	*RMSE*	1.13	1.07	1.38	1.35	1.26	1.21
	RE	0.70	0.66	0.99	0.97	0.84	0.81
10 days	*RMSE*	0.83	0.79	1.37	1.28	1.10	1.04
	RE	0.65	0.62	1.20	1.13	0.93	0.88
Storm events	*RMSE*	1.41	1.25	1.49	1.24	1.45	1.25
	RE	0.61	0.54	0.64	0.53	0.63	0.54

Table 7.5 Performance of the model for the estimation of TP concentration shown by the values of selected evaluation criteria

Survey/data types	Evaluation Criteria	Stations					
		St.37		St.38		St.(37+38)	
		Normal	Modified	Normal	Modified	Normal	Modified
All data	*RMSE*	0.03	0.03	0.04	0.04	0.04	0.03
	RE	0.66	0.59	0.92	0.83	0.79	0.71
Alternate days	*RMSE*	0.03	0.03	0.05	0.04	0.04	0.04
	RE	1.18	1.14	1.18	1.13	1.18	1.14
10 days	*RMSE*	0.03	0.03	0.05	0.05	0.04	0.04
	RE	1.05	1.00	1.41	1.34	1.23	1.17
Storm events	*RMSE*	0.02	0.02	0.04	0.03	0.03	0.02
	RE	0.36	0.24	0.63	0.51	0.50	0.38

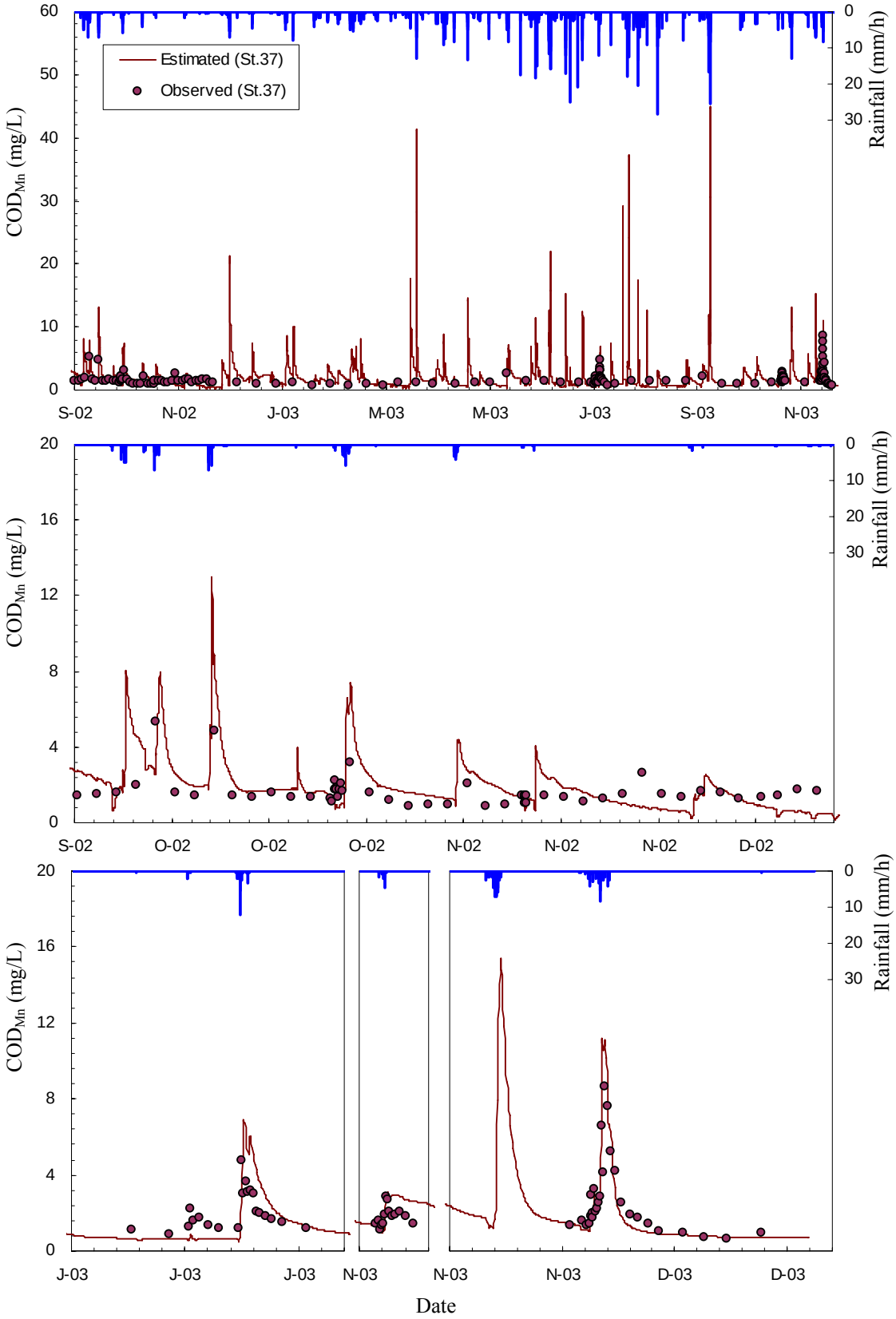


Figure 7.7 Estimated and observed COD_{Mn} at St. 37

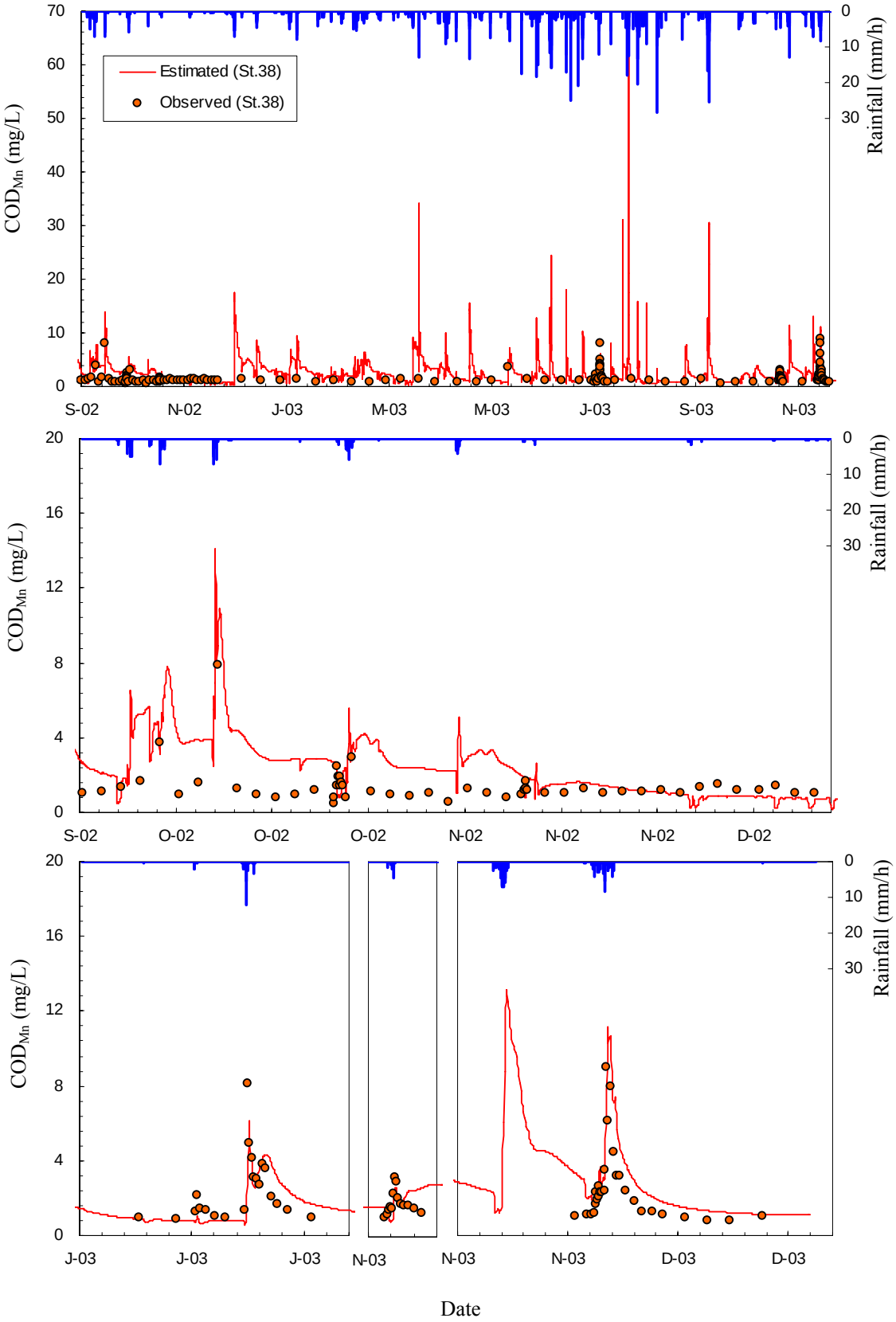


Figure 7.8 Estimated and observed COD_{Mn} at St. 38

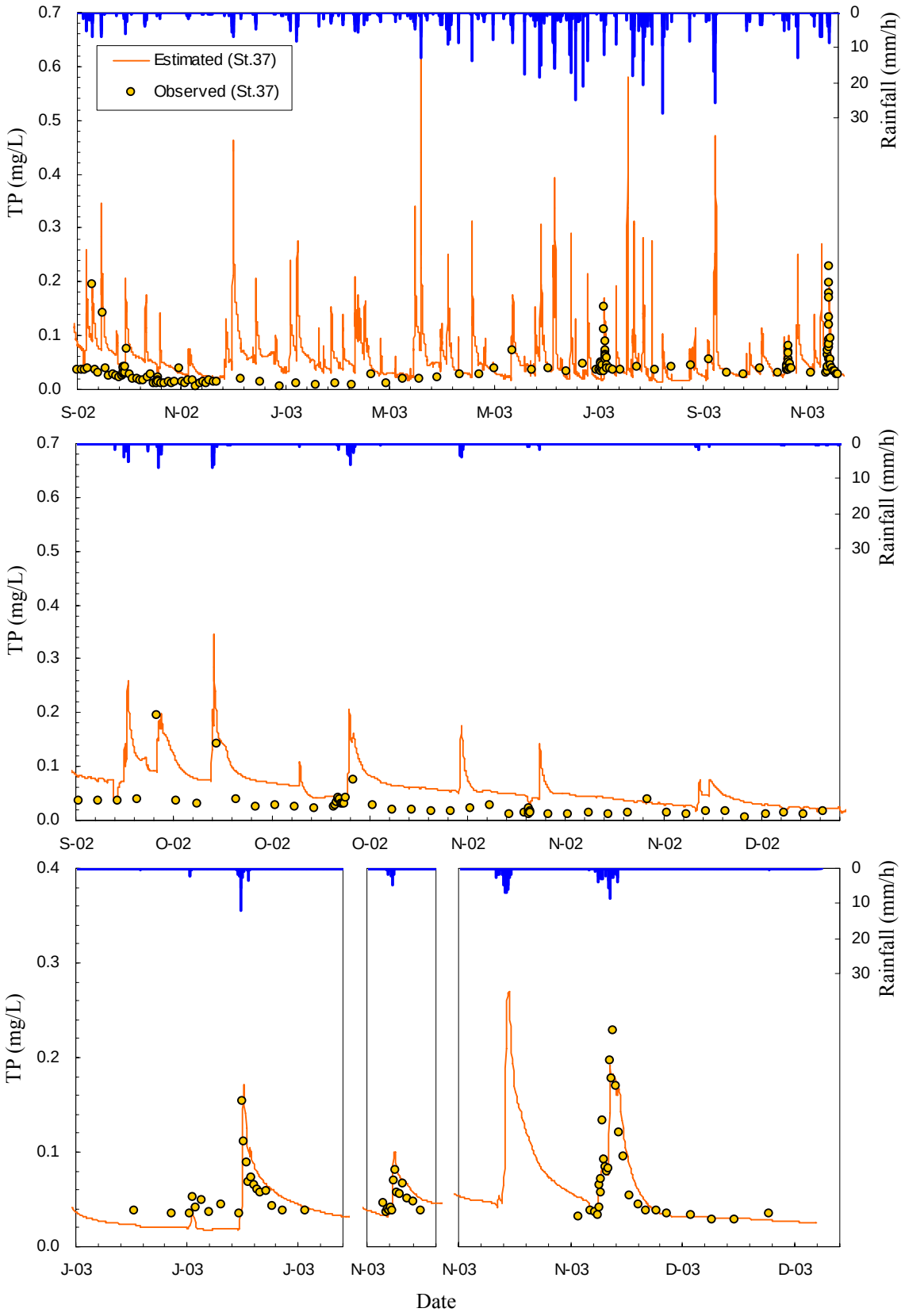


Figure 7.9 Estimated and observed TP at St. 37

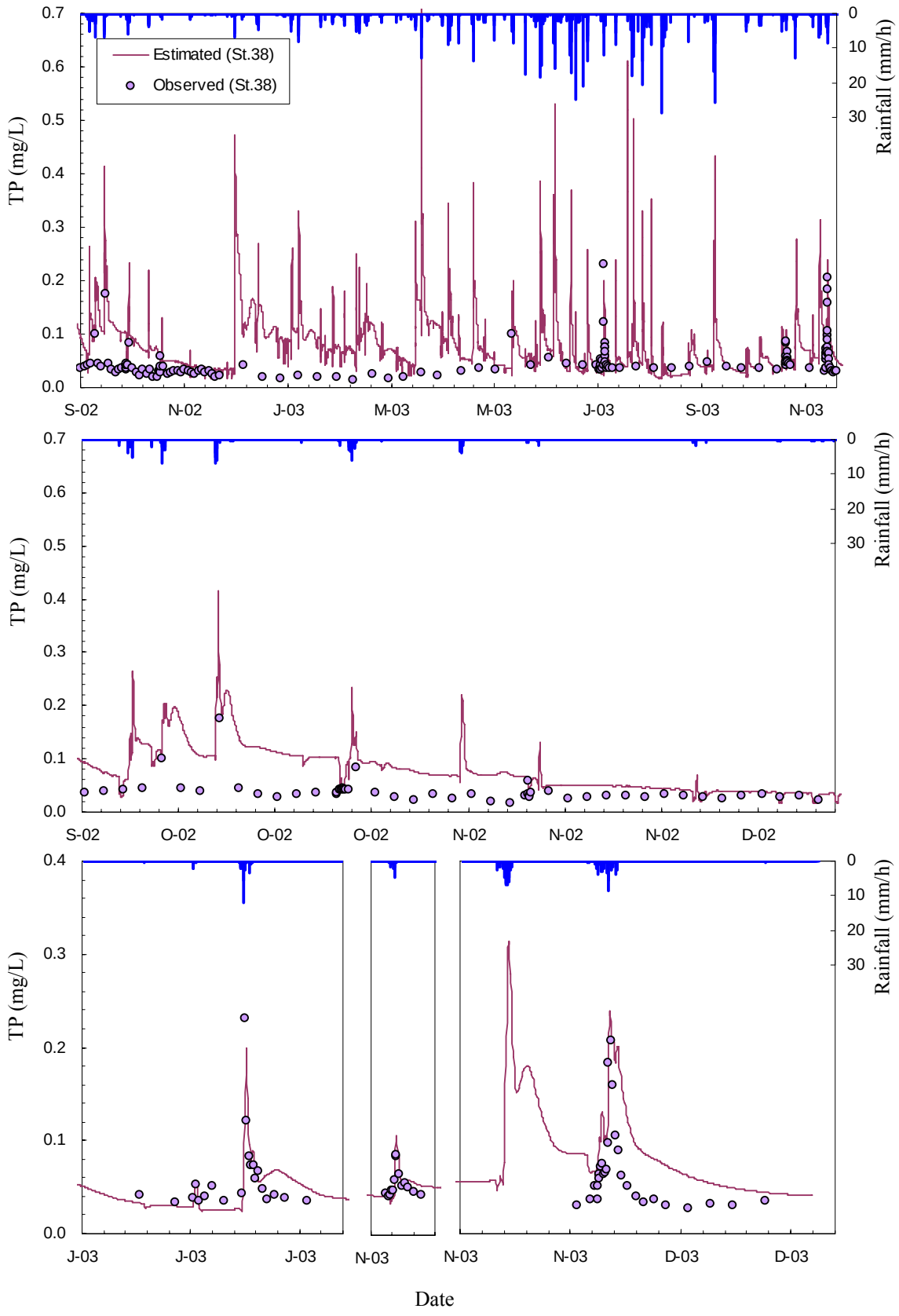


Figure 7.10 Estimated and observed TP at St. 38

However, *RE* was relatively lower than other cases, which could mean that the model estimation was quite satisfactory during the peak runoff periods. Comparison between observed and estimated results of flow rates at different spatial and temporal conditions indicated that model was able to estimate the river runoff pattern at different conditions. Therefore, the estimated runoff could be utilized to estimate the river water quality.

Figure 7.3b and 7.3c showed the comparison between estimated and observed COD_{Mn} and TP concentration at 39 sampling stations respectively. In comparison with estimated flow rate at 39 sampling stations, higher correlation coefficient could not be observed. However, the estimated concentration did not deviate so much. The lower correlation coefficient, $r < 0.1$ for COD_{Mn} and $r = 0.26$ for TP, might be due to the specific factors that may cause small variation in the production of the loadings in each grid. In addition to that, the observed data only represented specific sampling date condition, so the estimation was considered only for reference purpose.

For better understanding, we compared the estimated concentration of COD_{Mn} and TP at St.37 and 38 for the whole simulation period. **Figure 7.7 ~ 7.10** show the observed and estimated concentration of selected WQIs for both stations, and **Table 7.4** and **Table 7.5** showed the values of selected evaluation criteria. *RMSE* for overall data (St.37 + St38) was 1.31 mg/L for COD_{Mn} (*RE*: 0.74) and 0.04 mg/L (*RE*: 0.79) for TP. Although observed and estimated pattern were quite similar at both stations, estimation at St. 37 showed close correspondence with the observed data than St.38. It could be due to the land cover differences, in which sub-basin of St.38 was more urbanized and have several agriculture fields. Majority of the differences occurred in the alternate days or 10 days data, which showed higher *RE*. For the peak periods, estimation pattern were similar for both stations showing $RE < 1$ in most of the cases. In all cases (including flow rates estimation), modified method showed relatively lower values of evaluation criteria than normal method, which indicate that few hours' differences in the observed and estimated results could significantly affect the values of evaluation criteria.

The comparison between observed and estimated results indicated that the model could estimate both river flow rates and the concentration of pollutants at different conditions quite satisfactorily. However, distributed models were reported to be sensitive to the scale of its grid, so further evaluation of the model estimation at different grid sizes would be required.

7.3.3 Effect of grid size on important watershed attributes

Before the comparison of estimated results among different grid sizes, it was necessary to examine the differences in the spatial attributes because of grid resizing. Its effects on the area and population were already adjusted, so they were not discussed anymore. **Table 7.6** shows the proportion and average values of different watershed variables at different grid sizes. It was obvious that with increasing grid size the number of cells will decrease, including channel cells. However, proportion of channel grids increased with increasing grid size. In case of 50m grid size, the channel grids accounted only 15.6% of total grids, but it was more than 42% for 500m grid size. Channels are assumed as features showing rapid runoff so the increased proportion of channel cells at bigger grid sizes could have significant influence on the estimated runoff **Figure 7.11** shows the average number of cells required to travel to the outlet and channel cells. It could be noticed that with increasing grid size the average number cells to reach the outlet was decreasing following nearly proportionate pattern. Its effect could be on the peak flow rate due to the rapid movement of the water from upstream to downstream.

Slope and critical outlet height for surface runoff (S_{sf}) were two essential parameters that could show wider variability inside the watershed than other categorical variables. They were also directly linked with the runoff process in the model, so it was essential to assess their distribution at different grid sizes. **Figure 7.12** shows the distribution of slope and S_{sf} at different percentiles. Slope decreased rapidly with increasing grid size, which was due to the increase in the grid area after resizing to bigger grid sizes. At 10% distribu-

Table 7.6 Effect of grid resizing on different basin attributes

Grid size (m)	Total grids	Channel grids	Land cover (%)			Soil profile (%)			Average	
			Forests	Urban (vegetations)	Urban area	Shallow	Medium	Deep	*Lai*	T_c
50	54005	8446	86.9	9.1	3.9	62.6	29.8	7.6	1.84	0.608
100	13562	2839	85.5	11.6	3.0	65.2	28.3	6.5	1.84	0.609
150	6161	1546	85.3	11.2	3.5	62.7	30.2	7.1	1.83	0.609
200	3487	981	85.3	11.1	3.6	63.4	29.9	6.7	1.82	0.609
250	2279	718	85.6	10.8	3.6	60.8	31.3	7.9	1.81	0.609
300	1589	519	85.5	10.6	3.8	61.7	30.1	8.1	1.81	0.607
350	1177	423	85.9	9.9	4.2	61.8	30.4	7.8	1.81	0.608
400	910	346	86.2	10.3	3.5	61.5	30.0	8.5	1.81	0.607
450	696	303	86.2	10.5	3.3	62.2	30.2	7.6	1.81	0.607
500	594	255	85.4	10.1	4.5	61.3	29.1	9.6	1.79	0.606

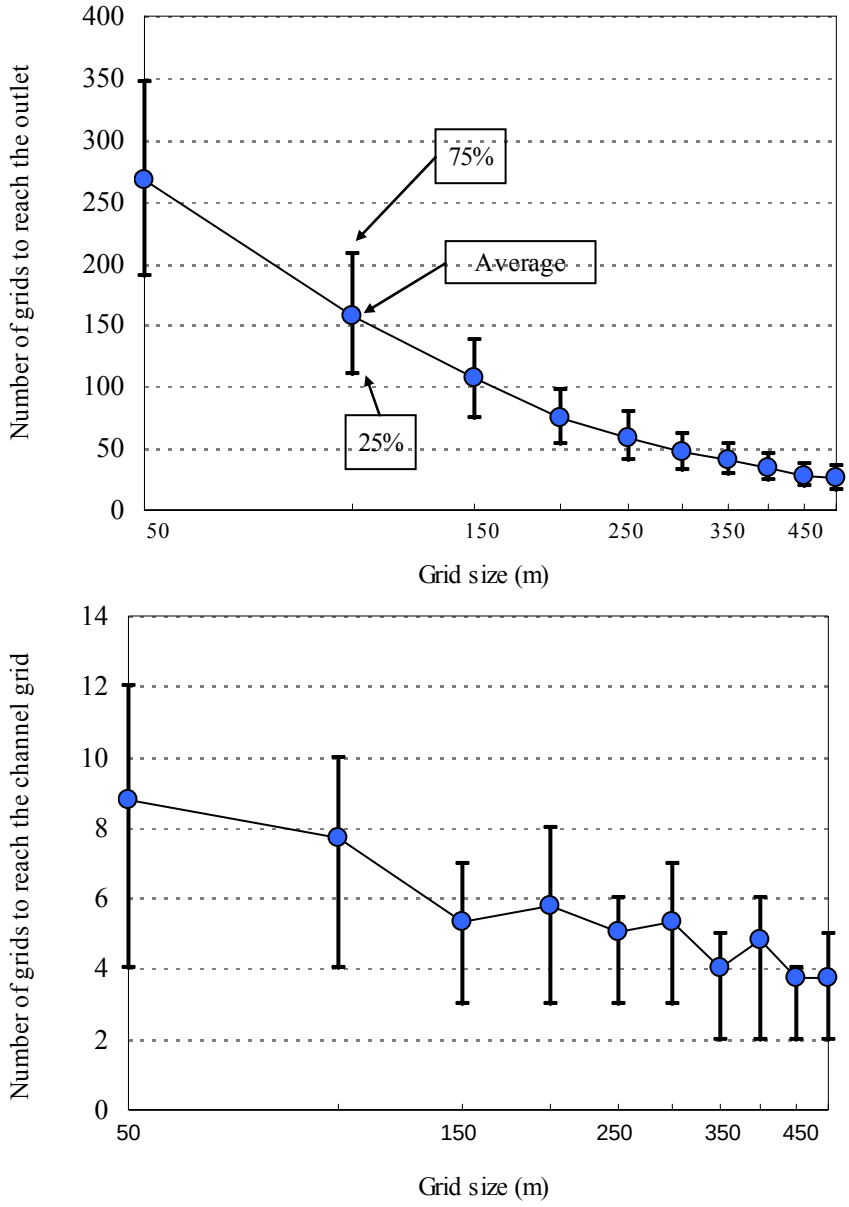


Figure 7.11 Effects of grid resizing on the average number of grids draining to outlet and any channel grids

tion point, the slope for 50m grid size was 0.035 but it decreased to 0.009 at 500m grid size. At 90% distribution point, the slope was nearly 0.8 for 50m grid size but it reduced to 0.21 at 500m grid size. Since slope is directly proportional to surface and sub-surface runoff, so decreasing slope would certainly increase the residence time of the water in the upper tank. In case of S_{sf}, initially it decreased rapidly until 200m grid size, however for 250 to 500m grid size its value was quite similar. S_{sf} value was mainly influenced by upslope contributing area. At bigger grid size, relatively few numbers of upstream grids but with

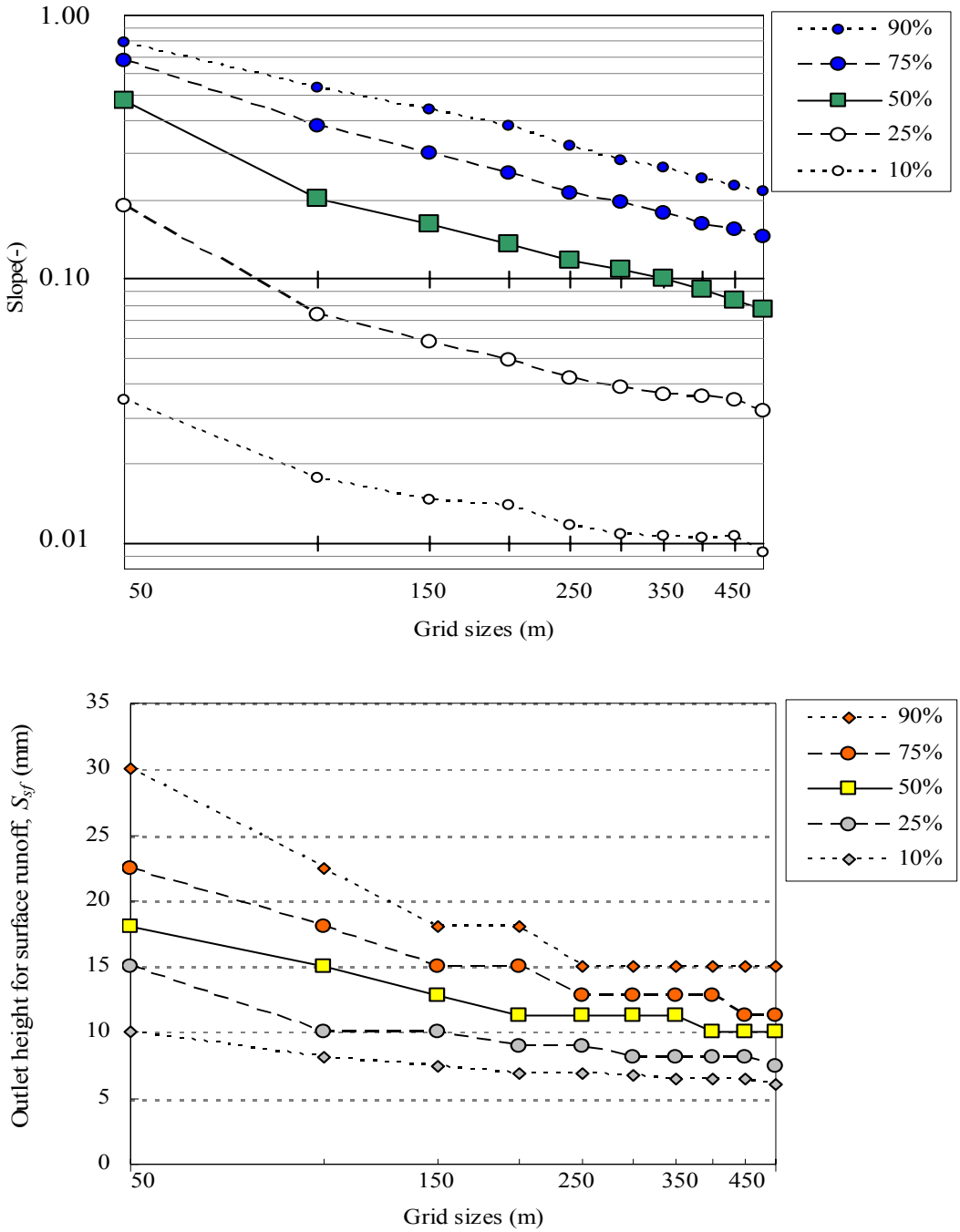


Figure 7.12 Distribution pattern of slope and S_{sf} at different grid sizes

increased area could make its values similar. Because of lower S_{sf} value at larger grid sizes, it could be possible that overland runoff generation by the model would initiate even after smaller rainfall events.

Land cover and geology (or soil profile depth) were two main categorical data types used in the model. Due to the rather homogenous land cover, significant differences could not be observed (Table 7.6). In case of land cover, only marginal differences were seen for

all categories. Forest proportion slightly decreased at 500m grid size, but remaining minor categories were slightly increased. Although increase in the minor categories was insignificant in comparison with the watershed area, the percentage increase by the category was quite significant. Soil profile category also showed similar trend, in which the proportion of deep soil profile increased slightly at bigger grid size. For other variables, such as *Lai* and T_c, changing grid size did not cause significant differences on their average values.

Smaller changes observed on the major land cover types or soil profile categories at different grid size could mean that their effects on the model estimation could be very limited. However, with increasing grid size, increasing proportion of channel grids, less number of grids to travel, and smaller values of S_{sf} could aggravate the runoff rate. Decreasing trend of slope with increasing grid size could mean that it could have inverse effect on the runoff rate. Therefore, number of grids, S_{sf} and slope together could exert balancing effect on the river runoff with increasing grid size.

7.3.4 Model estimation at different grid sizes

Simulation by using fixed parameter values:

It was already shown that 50m grid size showed satisfactory estimation for both flow rates and selected WQIs. Therefore, it was reasonable that the estimated hydrographs and pollutographs could be used as a basis to evaluate the estimated results of other grid sizes (Eq. 7.5). For this analysis, peak flow rate of a storm event survey (2003/11/28-2003/12/6) was used to visually analyze the effect of grid sizes on the estimated hydrographs (or pollutographs), but evaluation criteria (*RMSE* and *RE*) were also applied to assess the overall differences in the estimated values caused by the model for the whole simulation period. **Figure 7.13** showed the estimated flow rates and COD_{Mn} by the model at St.37 at different grid sizes by using the same parameters used in the case of 50m grid size (Table 7.1 and Table 7.2). Peak flow rate showed increasing pattern with the increasing grid sizes. Larger proportion of channel grids could have caused higher flow rates at bigger grid size. It meant that effect of slope was lesser than number of grids to reach the channel and total number of channel grids. For the COD_{Mn}, the concentration was quite lower at bigger grid size as compared with 50m. Higher runoff as compared with the constant production rates of the COD_{Mn} could have caused lower concentration. However, intermediate grid sizes

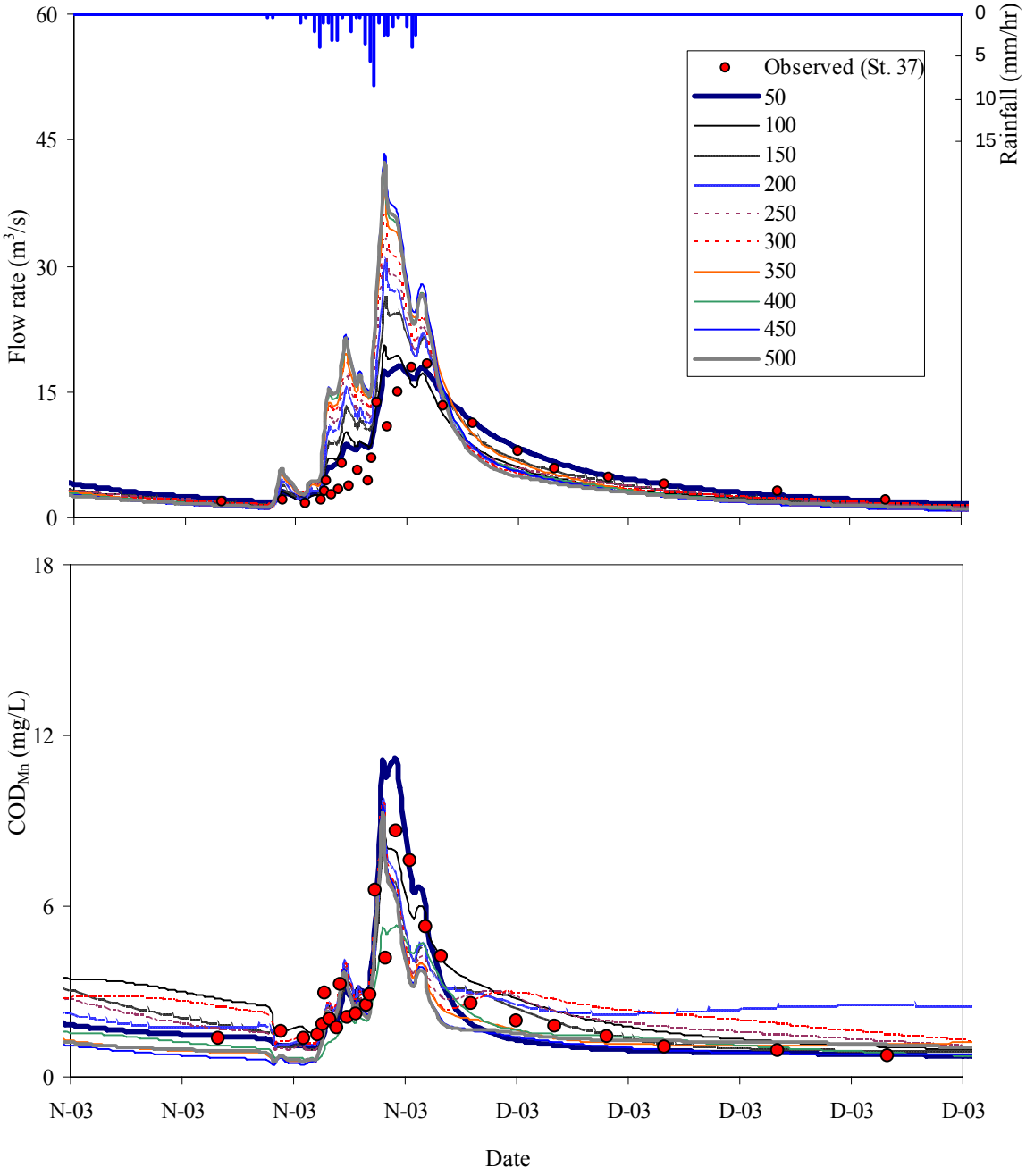


Figure 7.13 Estimated flow rates and COD_{Mn} concentration by the model at different grid sizes with same values of parameters

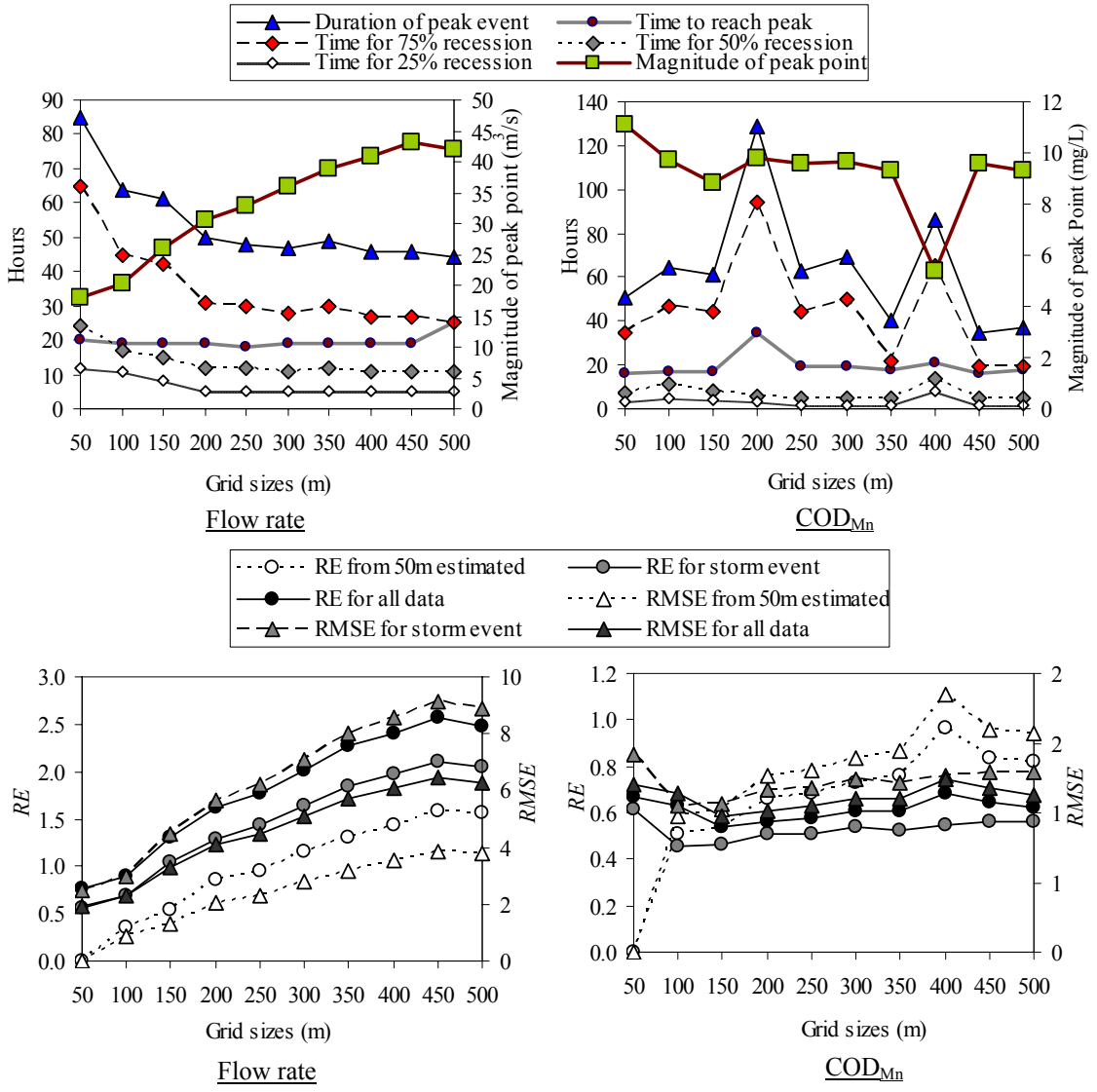


Figure 7.14 Comparison of a peak event estimated by the model for different grid sizes with same values of parameters

showed quite different patterns which could not be easily identified. It could mean that variation of runoff condition at different grid size could cause such differences on the estimated concentration of COD_{Mn}. **Figure 7.14** shows duration of peak runoff period and the magnitude of the highest peak points for flow rate and COD_{Mn}. Similarly, same figure also showed the evaluation criteria used for different grids sizes. Magnitude at peak flow rate point was nearly two times higher than 50m grid size. *RMSE* and *RE* also increased with increasing grid size. However, the duration of peak events as well as time of recession of the peak of flow rates was higher for 50m grid size than bigger grid sizes. It could mean that changing grid size not only affected the magnitude of peak runoff but also the shape of the hydrographs. For COD_{Mn}, the patterns at different grid sizes could not be observed clearly. Except some grid sizes, both magnitude and duration of peak concentration period did showed any variation. Both *RMSE* and *RE* only showed slightly increasing pattern at bigger grid size, which could be considered insignificant.

Assessment of the sensitivity of key parameters on the estimated results:

It was necessary to examine the model sensitivity to different parameters before assessing the actual effects of grid size on the values of model parameters. The assessment was important in that the parameters could be identified in terms of their effects on the model estimation. **Figure 7.15** showed the model estimation at different values of key parameters that were usually modified during the fitting with observed data. As mentioned earlier, the values of each parameter were changed such that values of other parameters (Table 7.1 and Table 7.2) were kept unchanged. The f_1 showed the increasing of the peaks by increasing its values, but it also caused widening of peaks while without changing much in the base flow periods. The f_0 also showed similar pattern but the effect on the magnitude of the peaks appeared smaller. It could mean that its effect on peak runoff was lesser. In case of f_r both magnitude and shape of the peaks showed remarkable changes. In case of g_1, increasing its value caused homogenous increase or shifting in the whole hydrograph. It was usually related with lower tank that conceptualizes groundwater flow, so its impact on the peak runoff should be lesser as compared with base flow periods. Since major interest in the analysis was on the peak periods, so f_1 and f_r were selected for the calibration at different grid sizes. For the water quality, there were only two parameters to adjust, namely, a and c. Model estimation by changing the values of both parameters was quite

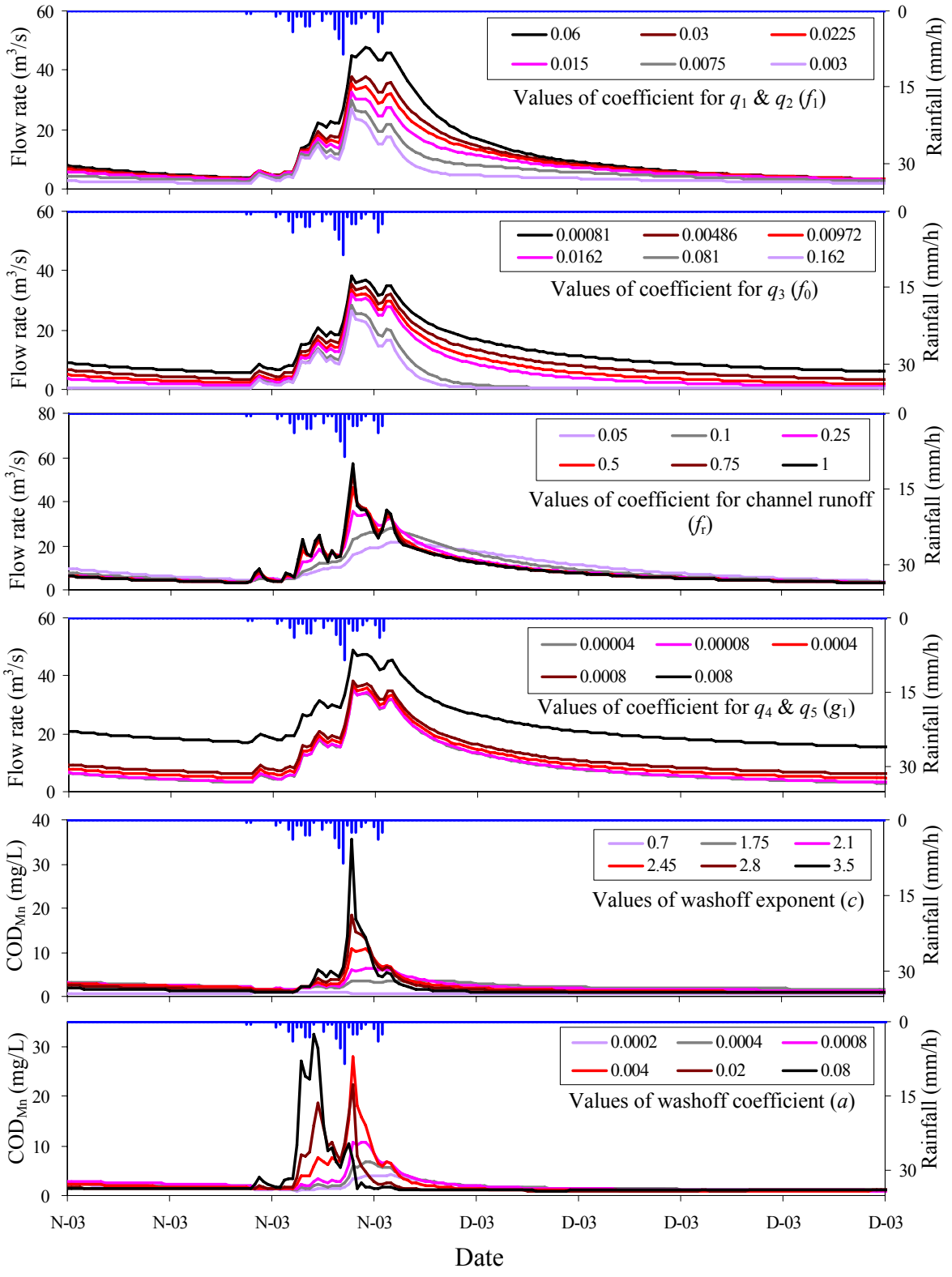


Figure 7.15 Model estimation at different values of selected parameters

noticeable. Increasing of c mainly caused the higher magnitude of the peaks but narrowing of the pollutograph. As c has exponential effect on the estimated runoff, so after adjusting

the effect of grid size on estimated runoff, it effect should not appear significant on the estimated concentration. In case of *a*, its effect was mainly seen on magnitude, shape, and timing of peak initiation, in which its higher values even caused the initiation of peaks for relatively smaller rainfall event (or runoff peaks). The effect of *a* was also seen on the timing of the peak initiation, so it was selected for the calibration.

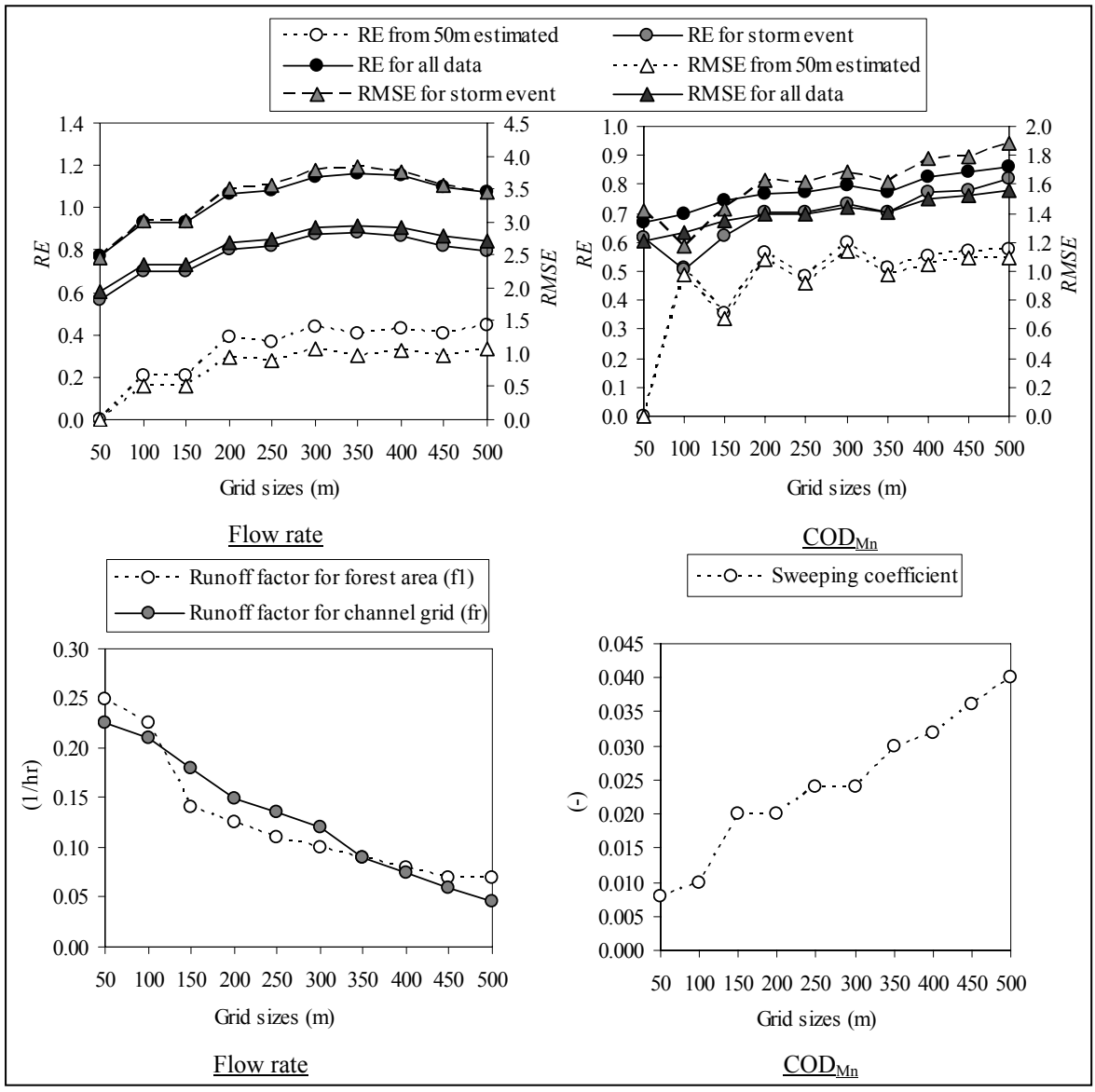


Figure 7.16 Values of evaluation criteria for different grid sizes (above) obtained after adjusting the values of key parameters (below)

Simulation by using changing values of key parameters:

The values of key parameters (outlet coefficients of upper tank, f_1, and channel grids, f_r and sweeping coefficient, a) were changed as long as the error became minimum with respect to the estimated value of 50 m grid size. **Figure 7.16** shows the *RMSE* and *RE* (above portion) obtained from the simulation after adjusting the values of the parameters (below portion) for each grid size. The runoff parameter was given for the forest land cover. It could be observed that, even after adjusting by calibration, some errors were still remained. It was mainly due to the shape of the peak period hydrograph.

In case of flow rate estimation, both f_1 and f_r showed remarkable decreasing trend with increasing grid size. In fact, it was necessary to reduce the values of the coefficient to decrease the magnitude of peak periods. However, in case of COD_{Mn}, it was necessary to increase the values sweeping coefficient to account for the changes in the shape of hydrograph obtained after adjusting. The observed patterns of the values of selected parameters indicated that it was possible to adjust the values in an identifiable pattern. In comparison with the extent of effects on the watershed attributes, grid size effects could be normalized only by changing the values of key parameters at relatively narrow range (0.05 to 0.25 for runoff coefficients and 0.005 to 0.04 for sweeping coefficient). Therefore, it may be inferred that grid size and values of parameters could be related in some systematic pattern and could be adjusted by small changes in the values of some critical parameters.

7.3.5 Model estimation at different scenario

Effects of rainfall condition on the model estimation:

The basic 50m grid size was used to assess the model at different scenario conditions because the distributed data were derived directly from the available sources. **Figure 7.18** shows the estimated results in terms of cumulative flow rate and COD_{Mn} at selected annual rainfall events [(maximum (2200 mm/yr), year 2003 (1813 mm/yr), and minimum (800 mm/yr)] at St.37. It was obvious to observe the clear differences by different rainfall conditions, but cumulative runoff at maximum rainfall did not showed much difference with that of year 2003 even with huge difference of 387 mm of rainfall. Therefore, few hundreds mm of rainfall differences may not be significant to show remarkable differences in the estimated result. The average runoff rates were 2.41, 2.32, and 0.59 m^3/s for maximum, 2003, and minimum rainfall periods respectively. Especially, at minimum

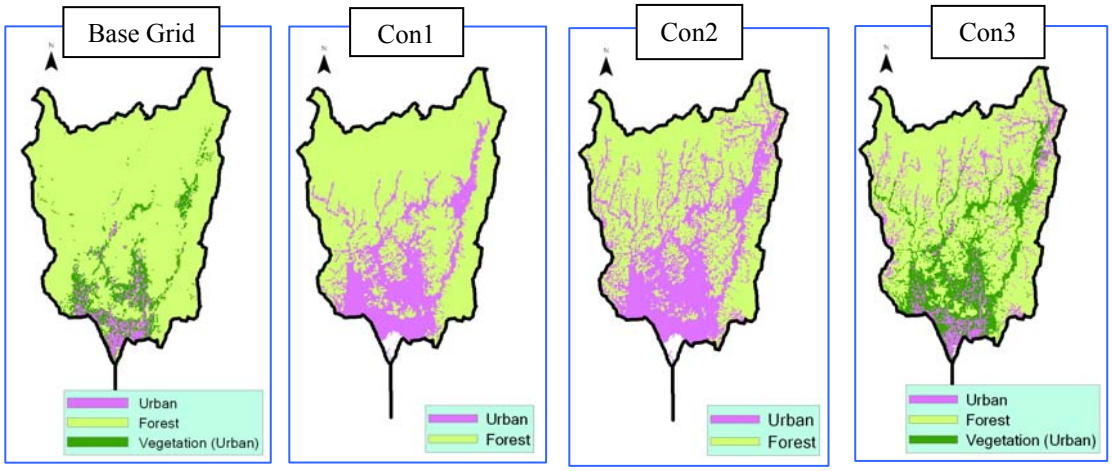


Figure 7.17 Land cover maps at different scenario conditions

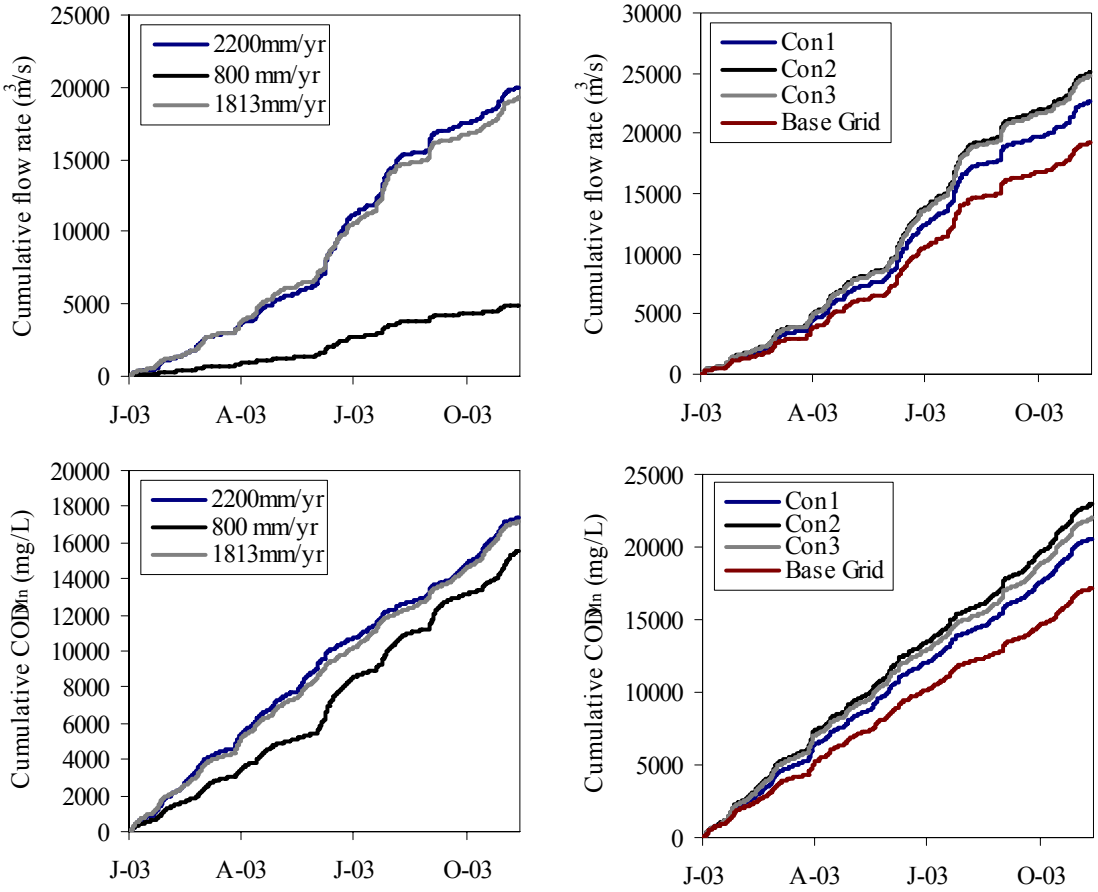


Figure 7.18 Cumulative of estimated flow rates and COD_{Mn} concentration at extreme rainfall events year and at different land cover scenarios.

rainfall of 800 mm/yr, the average runoff was quite lower. Compared between 2200 and 800 mm/yr of rainfall, the difference between the maximum peak event was quite huge (~ 30m^3/s). For COD_{Mn}, the concentration differences were not remarkable in comparison with the flow rate. Lower concentration of COD_{Mn} at 800 mm/yr rainfall in comparison with other rainfall conditions could be due the decrease in the erosivity of flowing water. The average concentrations were 2.10, 2.07, and 1.87 mg/L for maximum, 2003, and minimum rainfall periods respectively. The results indicated that the change in runoff conditions could have relatively higher impact on resultant runoff than water quality.

Effects of the change in land cover on the model estimation:

The effects of land cover change on the model estimation were evaluated by considering the three conditions as explained earlier. **Figure 7.17** shows land cover maps of the original and modified conditions. In the base map, the proportion of urban, forests and vegetation urban were 4.15%, 86.12% and 9.37% respectively. Similarly, the proportion for other land cover conditions was as Con1: 24.77%, 75.23%, and 0%, Con2: 36.82%, 63.18% and 0%, and Con3: 12.79%, 68.63%, and 18.59% of the total area. **Figure 7.18** shows the estimated results in terms of cumulative flow rate and COD_{Mn} at selected land cover scenarios. Small changes were seen on both the estimated flow rates and COD_{Mn} concentrations, but magnitude showed some proportionate change in the order of Con2> Con3> Con1> base grid size. Parameters for runoff and loadings production in both urban and agricultural areas in quite close range so the average flow rate and COD_{Mn} concentration difference between Con2 and Con3 was only 0.04 m^3/s and 0.14 mg/L respectively. Although the differences were quite small, the trend clearly indicates that models could simulate accordingly with the changes in the land cover proportions.

7.4 Summary

Three approaches were used for the model assessment. First was the comparison of the estimated results with the observed data, taken from different sampling stations. The second was the evaluation of model estimation at different grid sizes. The third was the model simulation at different scenario.

Initial assessment of the model at 39 sampling stations and for the daily river water depth showed quite satisfactory estimation of river flow rates for both spatial and temporal

conditions for the basic grid size (50 m). Similarly, estimated flow rates with respect to the different temporal surveys also gave reasonable fitting with the observed data showing lower *RMSE* and *RE*. On the basis of estimated runoff, simulation was performed to estimate the selected water quality indices, namely, COD_{Mn} and TP. Both water quality indices were well estimated by the model following the observed pattern, but more rigorous parameterization could further improve the estimation. However, estimated water quality indicated that the model could be developed further to simulate range of water quality indices after modifying certain parameters or modeling process.

Analysis of the effects of grid size on both basins' attributes and estimated results by the model were found to have some differences showing proportionate pattern. Especially the physical attributed such as number of grids, slope, and critical outlet heights showed remarkable differences at different grid sizes than categorical attributes. Assessment of the model estimation at different grid size showed wider variation for the estimated results. With increasing grid size, the estimated peak flow rate also followed increasing trend. Estimation of water quality was also affected by changing the grid size but clear pattern could not be understood from that. Individual assessment of the model for each grid size by changing essential parameters (f_l, f_r and a) showed that the model could be adjusted to attain targeted error range by following some proportionate pattern. It indicated that the values of selected parameters could be directly related with the estimation accuracy. It further indicates that consideration of the scale of spatial data in distributed modeling is quite important and need to be researched further.

The model estimation at different scenario indicated that few hundreds mm/yr of rainfall difference may not cause remarkable change in the river runoff as compared with the runoff simulated at driest and wettest year. It was seen that the changes in the estimated COD_{Mn} was quite lower in comparison with the extent of the change in the runoff rate. Even with the change in different land cover scenario, the estimated differences were not so higher. However, in all above cases, increasing trend in the runoff or concentration of COD_{Mn} was observed with the increase in either annual rainfall or decrease in the forest cover. The trend indicates that model could simulate the effects of changes in biophysical conditions inside the watershed.

CHAPTER VIII

CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusions

Watershed processes related with water quantity and quality are quite complex to understand. River systems are probably the most dynamic among the water resources that are directly interlinked with the watershed processes. In case of Japan, most of the watersheds are dominated by forest cover, which has a vital role in the conservation of both soil and water. Besides, these watersheds help to maintain the flow of the rivers in Japan, which constitute the dominant supplier of the fresh water accounting more than 70% of total demand. Therefore understanding the characteristics of these forested watersheds as well as their influence on the flow regime of river systems is vital for the management of both river water quantity and quality.

Modeling is widely used to understand the flow processes, to assess the impact of flooding or non-point source pollution, or for decision making by the watershed managers. However, even after the advent of distributed modeling concept, available models are quite complex in comparison with the available data to assign the values of parameters. Such condition often resulted in the smaller predictive performance causing uncertainty in the modeling process. In this study, a new distributed water quantity and quality model was developed to sort out some of issues related with distributed modeling. At first, relevant data were prepared in raster format that was required by the model (Chapter 3). Then water quantity and quality data representing different spatial and temporal conditions were collected and analyzed to understand the major variation pattern of the river water quantity and quality, which will be used latter for the verification of the model (Chapter 4). In the Chapter 5, distributed water quantity and quality model was developed especially focusing on the simple and less parameter intensive structure. Then remote sensing techniques and GIS were used to derive the values of major parameters of distributed model (Chapter 6). Finally, the applicability of the model to estimate the both river water quantity and quality was assessed by the comparison between the observed and estimated data, as well as

through the assessment of the model at different scale of the grid. In addition, the model was also used to evaluate two scenarios, by the change in rainfall patterns and land cover. Brief overview of the main results of each chapter is summarized hereafter.

Chapter III: Description of study area and database construction:

In this chapter relevant data (spatial, meteorological, population) were prepared and most of those data were converted into grid (or raster) format. From the preliminary assessment of the study area and its attributes, following aspects were understood:

-Study areas was dominated by the hilly topography, mostly covered by the forest in the upper part and urban areas in the lower part representing similar features with other watersheds in central Japan.

-Frequent and intense rainfall, wide variation in day-to-day as well as seasonal climatic conditions was another important characteristic of the area;

Chapter IV: Water quantity and quality surveys and their spatial and temporal variation analysis:

River water quantity and quality surveys were conducted at both spatial and temporal conditions in order to examine their variation pattern and for the verification of model estimation. The important findings related with this chapter could be briefly summarized as:

1) Rainfall-runoff pattern was quite rapid after the initiation of rainfall, in which the duration of peak event usually lasted less than two days;

2) Two types of water quality indices showing dilution effect (*e.g.*, IC, Cl^-, SO_4^{2-}, Na^+, K^+, Mg^{2+}, Ca^{2+}, Sr^{2+}, and Ba^{2+}), and sweeping effect (*e.g.*, T-COD_{Mn}, D-COD_{Mn}, DOC, SS, VSS, TP, DP, NH_4^+-N, Al) with the resultant runoff were identified from the analysis.;

4) Majority of WQIs covary together at different spatial and temporal conditions. Possibly the effects of acid rain, steep topography, and the biogeochemical processes in the subsurface of forest areas could be the main cause of covariance of majority of WQIs. The covariance of WQIs further indicates that the rainfall-runoff processes and the resulting water quality are closely interlinked with the biogeochemical processes in the forest areas;

Chapter V: Development of a distributed water quantity and quality model:

Distributed water quantity and quality model was developed which has its conceptual framework typical to the lumped tank model. The model was characterized by its simple conceptual framework, less parameter intensive, and familiar *MS-Excel* based graphical user interface.

Major emphasis was given to the uniqueness of the biophysical characteristics of the watershed by utilizing the available spatial information in the modeling, such as by the use of remote sensing and GIS. It was a significant improvement over the lumped type tank model that mainly relies on calibration and validation for the values of its parameters. The model was expected to be useful to estimate the river discharge and river water quality as an output from diffuse sources, especially, from the hilly and forested watersheds

Chapter VI: Application of remote sensing and geographic information system (GIS) in distributed model:

Remote sensing and GIS techniques were used to derive watershed attributes that were indispensable both during the design of the model and to assign the values of the distributed parameters. Three seasonal Landsat ETM+ images were processed to derive the land cover map of the study area including major forest types, such as deciduous, mixed, and evergreen.

Leaf area index (*Lai*) was determined with the help of land cover map and by the use of reflectance of two thematic bands (*TM*3 and *TM*4) of Landsat ETM+. *Lai* was required for the estimation of canopy water storage during the modeling process.

With the help of *TM*3 and *TM*4 bands, transpiration coefficient of green vegetation was determined, which was an essential requirement to estimate the evapotranspiration loss.

Evapotranspiration (*Et*) was estimated by FAO Penman-Monteith method, by which more than 4 mm/d during summer period and less than 2 mm/d of *Et* loss during winter was estimated.

Important topographic variables, such as slope, drainage direction, upslope contributing area, and outlet height for surface runoff were determined by the processing of DEM using GIS techniques.

Chapter VII: Application of the distributed water quantity and quality model in the study area:

This chapter focused on assessing the performance of the distributed water quantity and quality model. Initial assessment of the model at 39 sampling stations and for the daily river water depth showed quite satisfactory estimation of river flow rates for both spatial and temporal conditions. Estimated flow rates with respect to the different temporal survey gave reasonable fitting with the observed data showing lower root mean square error (*RMSE*) and relative error (*RE*).

On the basis of estimated runoff, simulation was performed to estimate the selected water quality indices, namely, COD_{Mn} and TP. Both water quality indices were well estimated by the model within the acceptable range especially at different runoff conditions. The estimated results indicated that model could be available to estimate other water quality indices after necessary modification.

The parameter values were then applied to simulate the model at different grid sizes. With increasing grid size, the peak flow rate estimation also followed increasing trend. It was possibly due to the increased proportion of channel grid and less number of grids to travel that caused the rapid accumulation of runoff volume. Estimation of water quality was also affect by changing the grid size but clear pattern could not be understood.

Then the model was individually assessed for each grid size by changing essential parameters to minimize the difference between the observed and estimated data. Values of parameters for the surface runoff required to be minimized to attain the desired accuracy. In this case, water quality indices were closer to the estimated values, but they were also adjusted to fit with the observed data. The effects of grid sizes on the model estimation showed that basin attributes at different scale grids could be proportionately related with the model estimation. Overall it was understood that simple assessment of distributed model based on the observed and estimated results may not give stable results when scale of distributed variables related with model parameters were changed. It could further indicate that the choice of scale should not be arbitrary rather they should be assessed before the actual application considering the spatial variability of the watershed.

Estimated results under extreme rainfall periods and different land cover scenarios were also evaluated. The distinct differences observed in both estimated water quantity and quality indicated that model could be utilized to assess the likely changes in future.

8.2 Recommendations

The satisfactory estimation of the model in the study area further implies that, it could also estimate successfully in other hilly and forested watershed after performing preliminary assessment of the values of model parameters. The model should be further improved to simulate range of water quality indices occurring in the natural rivers. In addition to the selection of simple distributed model to suit the available data and spatial variability, the study recommends prior analysis of any distributed models to ascertain that the magnitude of effects of grid size on the model performance and its parameters. Especially, the effects of grid scale on the model estimation should be evaluated using the data derived from primary sources rather than through resizing of the grids.

In terms of modeling approach, it was recommended that further enhancement of the methods of spatial processing would be required, especially, to derive the meaningful variables from the remote sensing and GIS techniques. Improving the understanding of distributed hydrological as well as water quality variables derived from these techniques could not only help to minimize uncertainty in the estimation but also increase the possibility of applying more physically based distributed models.

REFERENCES

Allen, R. G., Pereira, L. S., Raes, D., Smith, M. (1998). *Crop Evapotranspiration - Guidelines for Computing Crop Water Requirements.* FAO Irrigation and drainage paper 56, FAO.

Anazawa, K. and Ohmori, H. (2001). "Chemistry of surface water at a volcanic summit area, Norikura, Central Japan: multivariate statistical approach", *Chemosphere*, **45**, 807-816

Andersen, J., Dybkjaer, G., Jensen, K.H., Refsgaard, J.C. and Rasmussen,K. (2002). "Use of remotely sensed precipitation and leaf area index in a distributed hydrological model", *J. Hydro*., **264**, 34-50

Anderton, S., Latron, J. and Gallart, F. (2002). "Sensitivity analysis and multi-response, multi-criteria evaluation of a physically based distributed model", *Hydro. Process.*, **16**, 333-353.

Arheimer, B. and Olsson, J. (2003). *Integration and Coupling of Hydrological Models with Water Quality Models: Applications in Europe*. World Meteorological Organisation, WMO Technical reports in hydrology and water resources, No.75. WMO/TD-No.1174. Geneva.

Armstrong, R.N. and Lawrence, W. M. (2003). "Topographic parameterization in continental hydrology: a study in scale", *Hydro. Process.*, **17**, 3763-3781.

Arnold, J.G., Srinivasan, R., Muttiah, R.S. and Williams, J.R. (1998). "Large area hydrologic modeling and assessment part I: model development", *J. American Wat. Res. Association*, **34**(1), 73-89.

Asano, Y. and Uchida, T. (2005). "Quantifying the role of forest soil and bedrock in the acid neutralization of surface water in steep hillslope", *Environmental Pollution*, **133**, 467-480

Asano, Y., Uchida, T. and Ohte, N. (2003). "Hydrologic and geochemical influences on the dissolved silica concentration in natural water in a steep headwater catchment", *Geochimica et Cosmoshimica Acta.*, **67**(11), 1973-1989

Asselman, N.E.M. (2000). "Fitting and interpretation of sediment rating curves", *J. of Hydrology.*, **234**, 228-248.

Bashford, K.E., Beven, K.J. and Young, P.C. (2002). "Observational data and scale -dependent parameterizations: explorations using virtual hydrological reality", *Hydro. Process.*, **16**, 293-312.

Bengraine, K. and Marhaba, T.F. (2003). "Using principle component analysis to monitor spatial and temporal changes in water quality", *J. Hazardous Materials*, **B100**, 179-195.

Beven, K. J. (2001a). *Rainfall-Runoff Modelling: The Primer*. J. Wiley and S.

Beven, K. J. (2001b). "How far can we go in distributed hydrological modelling?", *Hydrology and Earth System Sciences*, **5**(1), 1-12

Beven, K. J. and Krikby, M. (1979). "A physically based variable contributing area model of catchment hydrology", *Hydrological Sciences Bulletin*, **24**, 43-69.

Beven, K. J. (1995). "Linking parameters across scales: subgrid parameterization and scale dependent hydrological models", *Hydro. Process.*, **9**, 507-525.

Biftu, G. F., and Gan, T. Y. (2001). "Semi-distributed, physically based, hydrologic modeling of the Paddle River Basin, Alberta, using remotely sensed data", *J. Hydro.*, **244**, 137-156.

Bloschl, G. (2005). "Rainfall-runoff modeling of ungauged catchments", In: M.G. Anderson (Ed), *Encyclopedia of Hydrological Sciences*, **Chapter 133**, J. Wiley and S.

Bruneau, P., Gascuel-Odoux, C., Robin, P., Merot, P. and Beven, K. J. (1995). "Sensitivity to space and time resolution of a hydrological model using digital elevation data", *Hydro. Process.*, **9**, 69-81.

Burrough, P. A. (1998). "Dynamic modeling and geocomputation", In: P. A. Longley (Edt.), *Geocomputation: A Primer*, J. Wiley and S., 165-191.

Chapman, D. and Kimstach, V. (1996). "Selection of water quality variables", In: Chapman D., (edt), *Water Quality Assessments-A Guide to Use of Biota, Sediments and Water in Environmental Monitoring*, 2nd edn., UNESCO/WHO/UNEP

Chapra, S. C. (1997). *Surface Water-Quality Modeling*, McGrew-Hill Pbs.

Chapra, S. C. and Pelletier, G. (2003). QUAL2K: A Modeling Framework for Simulating River and Stream Water Quality: Documentation and Users Manual, Civil and Environmental Engineering Dept., Tufts Univ., Medford, MA..

Chen, J., Barry, J., and Adams, J. (2006). "Analytical urban storm water quality models based on pollutant buildup and washoff processes", *J. Environ. Engg.*, **132** (10),

1314-1330.

Chen, J.M., Chen, X., Ju, W. and Geng, X. (2005). "Distributed hydrological model for mapping evapotranspiration using remote sensing inputs", *J. Hydro*., **305**, 15-39

Choudhury, B. J., Ahmed, N. U., Idso, S. B., Reginato, R. J., and Daughtry, C. S. T. (1994). "Relations between evaporation coefficients and vegetation indices studied by model simulations", *Remote Sens. Environ.,* **50**, 1-17.

Curran, P. J., Milton, E. J., Atkinson, P. M. and Foody, G. M. (1998). "Remote sensing: from data to understanding", In: P. A. Longley (Edt.), *Geocomputation: A Primer*, J. Wiley and S., 32-59.

Demayo, A. and Steel, A. (1996). "Data handling and presentation", In: Chapman D., (edt), *Water Quality Assessments-A Guide to Use of Biota, Sediments and Water in Environmental Monitoring*, 2nd edn., UNESCO/WHO/UNEP

DHI (2004).*The MIKE SHE technical reference*. *Working with MIKE SHE.* Danish Hydraulic Institute.

Dixon, W. and Chiswell, B. (1996). "Review of aquatic monitoring program design", *Water Research*, **30**, 1935-1948

Droogers, P., and Kite, G. (2002). "Remotely sensed data used for modelling at different hydrological scales", *Hydro. Process.*, **16**, 1543-1556.

Dymond, J.R. and Shepherd, J.D. (2004). "The spatial distribution of indigenous forest and its composition in the Wellington region, New Zealand, from ETM+ satellite imagery", *Remote Sens. Environ.,* **90**, 116-125

Evans, J.P. and Geerken, R. (2006), "Classifying rangeland vegetation type and coverage using a Fourier component based similarity measure", *Remote Sens. Environ.,* **105**, 1-8

Florinsky, I.V. (1998). "Combined analysis of digital terrain models and remotely sensed data in the landscape investigations", *Progress in Physical Geography*, **22**(1), 33-60.

Franchini, M., Wendling, J., Obled, C. and Todini, E. (1996). "Physical interpretation and sensitivity analysis of the TOPMODEL", *J. Hydro*, **175**, 293-338.

Frankenberger, J.R., Brooks, E.S., Walter, M.T., Walter, M.F. and Steenhuis, T.S. (1999). "A GIS-based variable source area hydrology model", *Hydro. Process.*, **13**, 805-822.

Freeze, R.A. and Harlan, R.L. (1969). "Blueprint for a physically based, digitally simulated hydrological response model", *J. Hydro*, **9**, 237-258.

Friedl, M. A., and Brodley, C. E. (1997). "Decision tree classification of land cover from remotely sensed data", *Remote Sens. Environ.,* **61**, 399-409

Grayson, R. and Bloschl, G. (2000). "Spatial modelling of catchment dynamics", In: R. Grayson and G. Bloschl (Edts), *Spatial Patterns in Catchment Hydrology*, Cambridge Univ. Press., 51-81.

GSI. (1997). *Digital Map 50m Grid (Elevation)*, Geographic Survey Institute, Japan

GSI. (1999). *Land Use Map of Kyoto Prefecture*, Geographic Survey Institute, Japan

Hardle, W. and Simar, L. (2003). *Applied Multivariate Statistical Analysis*, Springer.

Hashino, M., Yao, H. and Yoshida, H. (2002). "Studies and evaluations on interception processes during rainfall based on a tank model", *J. Hydro.*, **255**, 1-11.

Houser, P.R., Shuttleworth, W.J., Famiglietti, S. J., Gupta, H.V., Syed, K.H. and Goodrich, D.C. (1998). "Integration of soil moisture remote sensing and hydrologic modeling using data assimilation", *Water Res. Research*, **34** (12), 3405-3420.

Huete, A. R. (1988). "A soil adjusted vegetation index (SAVI)", *Remote Sens. Environ.,* **25**, 295-309

Ihara, H., Fujii, S., Nagare, H., Moriya, M., Songprasert, P., and Shimizu, Y. (2003). "Development of a distributed model for water quality and quantity in a river and its application for a river in Japan", *CD Proceedings of IWA Asia-Pacific Regional Conference*, Bangkok, Thailand.

Ikeda, H. and Miyanaga, Y. (1995). "Mechanism of acid-neutralization in two Japanese watersheds", *Water, Air and Soil Pollution*, **85**, 1867-1872

Ikeda, H. and Miyanaga, Y. (2001). "Hydrogeochemical conditions affecting acidification of stream water in mountainous watersheds", *Water, Air and Soil Pollution*, **130**, 1253-1258

Im, J. and Jensen, J.R. (2005). "A change detection model based on neighborhood correlation image analysis and decision tree classification" *Remote Sens. Environ***., 99,** 326-340

Jensen, J. R. (1983). "Biophysical remote sensing", *Annals of the Assoc. of American Geographers*, **73**(1), 111-132.

JARS (1999). *Remote Sensing Note*, Japan Association of Remote Sensing.

JMC (2007). Japan Meteorological Business Support Center, http://www.data.jma.go.jp

Jorgensen, S. E. (1989). "Principles of ecological modeling", In: W. J. Mitsch and S. E. Jorgensen (Edts.), *Ecological Engineering: an Introduction to Ecotechnology*, Wiley Pub.

Kato, T. (2005). "Development of a water quality tank model classified by land use for nitrogen load reduction scenarios", *Paddy Water Environ.,* **3**, 21-27

Kim, L. H., Zoh, K. D., Jeong, S., Kayhanian., and Stenstrom, M. K. (2006). "Estimating pollutant mass accumulation on highways during dry periods", *J. Environ. Engg.*, **132** (9), 985-993.

Kosugi, Y. and Katsuyama, M. (2007). "Evapotranspiration over a Japanese cypress forest. II. Comparasion of the eddy covariance and water budget methods", *J. Hydro.*, **334**, 305-311.

Kouwen, N., Soulis, E. D., Pietroniro, A., Donald, J., and Harrington, R. A. (1993). "Grouped response units for distributed hydrologic modeling", *J. Water Res. Planning. Mgmt.*, **119** (3), 289-305.

Krysanova, V., Muller-Wohlfeil, D.I. and Becker,A. (1998). "Development and test of a spatially distributed hydrological/water quality model for mesoscale watersheds", *Ecological Modelling*, **106**, 261-289.

Kuo, W. E., Steenhius, T. S., McCulloch, C. E., Mohler, C. L., Weinstein, D. A., DeGloria, S. D., and Swaney, D. P. (1999). "Effect of grid size on runoff and soil moisture for a variable-source-area hydrology model", *Water. Resour. Res.*, **35**(11), 3419-3428.

Kyoto Prefecture (1982).*Geological Map of Kyoto Prefecture,* 1: 50,000

Leavesley, G.H. and Stannard, L.G. (1990). "Application of remotely sensed data in a distributed-parameter watershed model", *Proc. Workshop on Application of Remote sensing in Hydrol.*, 47-68.

Lee, Y. H. and Singh, V. P. (1999). "Tank model using Kalman filter", *J. Hydro. Engg.*, **4**(4), 344-349.

Lee, Y. H. and Singh, V. P. (2005). "Tank model for sediment yield", *Water Resources Management*, **19**, 249-362.

Lillesand, T. M. and Kiefer, R. W. (2000). *Remote Sensing and Image Interpretation*, J. Wiley and S.

Maidment, D. R. (2002). *Arc Hydro: GIS for water resources*, ESRI.

Martin, J. L. and McCutcheon, S. C. (1999). *Hydrodynamics and Transport for Water Quality Modeling*, Lewis Pub.

Mclver, D.K. and Friedl, M.A. (2002), "Using prior probabilities in decision-tree classification of remotely sensed data", *Remote Sens. Environ.,* **81**, 253-261

Meiner, A. (1996). "Integration of GIS and a dynamic spatially distributed model for non-point source pollution management", *Water Sci. and Technol.*, **33**(3-4), 211-218.

Mather, P. M. (1999). *Computer Processing of Remotely-Sensed Images: An Introduction*, 2nd Edn., J. Wiley and S.

Milton, J. S. and Arnold, J. C. (1990). *Introduction to Probablity and Statistics: Principles and Applications for Engineering and the Computer Sciences*, 2nd Edn., McGrew-Hill.

Mitchell, M. J., Iwatsubo, G., Ohrui, K., and Nakagawa, Y. (1997). "Nitrogen saturation in Japanese forests: an evaluation", *Forest Eco. and Mgmt.*, **97**, 39-51

MOE (2003). *Quality of the environment in Japan 2003-white paper*, Ministry of Environment, Japan,

Morgan, R.P.C. (1995). *Soil erosion and conservation*, 2nd ed., Longman, London.

Murakami, S., Tsuboyama, Y., Shimizu, T., Fujieda, M. and Noguchi, S. (2000), " Variation of evapotranspiration with stand age and climate in a small Japanese forested catchment", *J. Hydro.*, **227**,144-127.

Muraoka, K. and Hirata, T. (1988). "Streamwater chemistry during rainfall events in a forested basin", *J. Hydro.*, **102**, 235-253

Nagatsuka, S. and Okazaki, M. (2005). Soil Map of Japan: Central Japan (1:1000,000)", In: Selvaradjou S K., Montanarella L., Spaargaren O., and Dent D., *EuDASM – Soil Maps of Asia DVD-ROM version*.

Nakamura, S., Takemura, M. and Kitano, Y. (1984). "Water quality and rock weathering in the upper reaches of the Tenryu river in connection with landslide disasters", *Jpn.J.Limnol.*, **45**(1), 13-25

Novotny, V. (2003). *Water Quality: Diffuse Pollution and Watershed Management*, 2nd edn. J. Wiley and S., 1 - 49

Oguchi, T., Saito, K., Kadomura, H. and Grossman, M. (2001). "Fluvial geomorphology and paleohydrology in Japan", *Geomorphology*, **39**, 3-19

Ohrui, K. and Mitchell, M.J. (1997) "Nitrogen saturation in Japanese forested watersheds", *Ecological Applications*, **7**(2), 391-401

Ohrui, K. and Mitchell, M.J. (1999). "Hydrological flow paths controlling stream chemistry in Japanese forested watersheds", *Hydrol. Process*., **13**, 877-888

Ohte, N., Mitchell, M.J., Shibata, H., Tokuchi, N., Toda, H. and Iwatsubo, G. (2001b). "Comparative evaluation of forest catchments in Japan and Northeastern United States", *Water, Air and Soil Pollution*, **130**, 649-654

Ohte, N., Tokuchi, N., Shibata, H., Tsujimura, M., Tanaka, T. and Mitchell, M.J. (2001a). "Hydrobiogeochemistry of forest ecosystems in Japan: major themes and research issues", *Hydrol. Process.*, **15**,1771-1789

Pal, M. and Mather, P.M. (2003). "An assessment of the effectiveness of decision tree methods for land cover classification" *Remote Sens. Environ***., 86,** 554-565

Pebesma, E.J., Switzer, P. and Loague, K. (2005). "Error analysis for the evaluation of model performance: rainfall-runoff event time series data", *Hydrol. Process.*, **19**,1529-1548

Perona, E., Bonilla, I. and Mateo, P. (1999). "Spatial and temporal changes in water quality in a Spanish river", *The Science of the Total Environment*, **241**, 75-90

Perrin, C., Michel, C. and Andreassian, V. (2001). "Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments", *J. Hydro*., **242**, 275-301

Refsgaard, J.C. (1996). "Terminology, Modeling protocol, and classification of hydrological model codes", In: M. B. Abbott and J. C. Refsgaard (Edts), *Distributed Hydrological Modeling*, Kluwer Aca.

Refsgaard, J.C. (1997). "Parameterisation, calibration and validation of distributed hydrological models", *J. Hydro*., **198**, 69-97.

Rogers, C.C.M., Beven, K.J., Morris, E.M. and Anderson, M.G. (1885). "Sensitivity analysis, calibration and predictive uncertainty of institute of hydrology distributed model", *J. Hydro*, **81**, 161-182.

Rossman, L. A. (2005). *Storm Water Management Model: User's Manual*, Version. 5, USEPA, 35-59.

Sato, K., Kishida, H. and Watanabe, T. (2004). "Nitrogen and phosphorous concentrations in rivers and lakes in Japan and their chronological change", *CD-Proceedings of IWA International Diffuse/Nonpoint Pollution Conference*, 24-29th Oct. 2004, Kyoto, Japan, **8**, 3-E-I-4

SBJ (2007). "*Japan Statistical Yearbook2007*", Statistics Bureau and Statistical Research

and Training Institute, Statistical Bureau Japan.

Schmugge, T.J., Kustas, W.P., Ritchie, J.C., Jackson, T.J. and Rango, A. (2002)., "Remote sensing in hydrology", *Adv. Water Resources*, **25**, 1367-1385

Schultz, G.A. (1993). "Hydrological modeling based on remote sensing information", *Adv. Space Res*., **13**(5), 149-166.

Schultz, G.A. (1988). "Remote sensing in hydrology", *J. Hydro*., **100**, 239-265.

Scott, A., and Clarke, R. (2000). "Multivariate techniques", In: Sparks, T. (ed.), *Statistics in Ecotoxicology*, J. Wiley and S.

Sellers, P. J, Randall, D. A., Collatz, G. J., Berry, J. A., Field, C. B., Dazlich, D.A., Zhang, C., Collelo, G. D. and Bounoua, L. (1996a). "A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: model formulation", *J. Climate*, **9**, 676-705.

Sellers, P. J, Los, S. O., Tucker, C. J., Justice, C. O., Dazlich, D.A., Collatz, G. J., and Randall, D. A. (1996b). "A revised land surface parameterization (SiB2) for atmospheric GCMs. Part II: the generation of global fields of terrestrial biophysical parameters from satellite data", *J. Climate*, **9**, 706-737.

Setiawan, B. I., Fukuda, T. and Nakano, Y. (2003). "Developing procedures for optimization of tank model's parameters", *Agricultural Engineering International: the CIGR J. Scientific Res. and Development*, **LW 01 006**, 1-13.

Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. and Kouimtzis, T. (2003). "Assessment of the surface water quality in Northern Greece", *Water Research*, **37**, 4119-4124.

Singh V. P. (1995). *Computer Models of Watershed Hydrology*, Water Resources Pub.

Singh, K., Malik, A. and Sinha, S. (2005). "Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques-a case study", *Analytica Chimica Acta*, **538**, 355-374

Singh, K., Malik, A., Mohan, D. and Sinha, S. (2004). "Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti river (India)-a case study", *Water Research*, **38**, 3980-3992

Smith, M. B., Seo, D.J., Koren, V.I., Reed, S.M., Zhang, Z., Duan, Q., Moreda, F. and Cong,S. (2004). "The distributed model intercomparison project (DMIP): motivation and experimental design", *J. Hydro*., **298**, 4-26

Sugawara, M. (1979). "Automatic calibration of the tank model", *Hydrolo. Sci. Bull*.,

24,357-388.

Sugawara, M. (1995). "Tank model", In: *Computer Models of Watershed Hydrology*, V. P. Singh (ed.), Water Resources Pub., 165-214.

Suzuki, T., Terakawa, H., and Matsura, T. (1996). "Development of physically-based distributed model for operational hydrological forecasting", *Civ. Engrg. Tec. Rep.*, Tsukuba, Japan, **38**(10), 26-31. (in Japanese)

Takahashi, M., Sakata, T. and Ishizuka, K. (2001). "Chemical characteristics and buffering capacity of surface soils in Japanese forests", *Water, Air and Soil Pollution*, **130**, 727-732

Tingsanchali, T. and Gautam, M. R. (2000). "Application of tank, NAM, ARMA and neural network models to flood forecasting", *Hydrol. Process.*, **14**, 2473-2487

Tsujimura, M., Onda, Y. and Ito, J. (2001). "Stream water chemistry in a steep headwater basin with high relief", *Hydrol. Process.*, **15**, 1847-1858

Vazquez, R. F., Feyen, J. and Refsgaard, J.C. (2002). "Effect of grid size on effective parameters and model performance of the MIKE-SHE code", *Hydrol. Process.*, **16**, 355-372

Vega, M., Pardo, R., Barrado, E. and Deban, L. (1998). "Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis", *Water Research*, **32**(12), 3581-3592

Vieux, B. E. and Needham, S. (1993). "Nonpoint-pollution model sensitivity to grid-cell size", *J. Water Res. Planning and Management*, **119**(2), 141-157.

Wagener, T., Lees, M.J. and Wheater H.S. (2002). "A toolkit for the development and application of parsimonious hydrological models", In: Singh, V.P. and Frevert, D.K. (eds.), *Mathematical Models of Large Watershed Hydrology*, Water Res. Pubs.

Wegehenkel, M., Jochheim, H. and Kersebaum, K.C. (2005). "The application of simple methods using remote sensing data for the regional validation of a semidistributed hydrological catchment model", *Physics and Chemistry of the Earth*, **30**, 575-587

White, A.F. and Blum, A.E. (1995). "Effects of climate on chemical weathering in watersheds", *Geochemica et Cosmochimica Acta*, **59**(9), 1729-1747

Wilson, E. H., and Sader, S. A. (2002). "Detection of forest harvest type using multiple dates of Landsat TM imagery", *Remote Sens. Environ.*, **80**, 385– 396.

Wolock, D. M. and Price, C.V. (1994). "Effects of digital elevation model map scale and

data resolution on topography-based watershed model", *Water. Resour. Res.*, **30**, 3041-3052.

Wood, E.F., Sivapalan, M., Beven, K. and Band, L. (1988). "Effects of spatial variability and scale with implications to hydrological modeling", *J. Hydro*, **102**, 29-47.

Wright, R. F., Roelofs, J. G. M., Bredemeier, M., Blanck, K., Boxman, A. W., Emmett, B. A., Gundersen, P., Hultberg, H., Kjonass, O. J., Moldan, F., Tietema, A., Breemen, N. and Dijk, H. F. G. (1995). "NITREX: responses of coniferous forest ecosystems to experimentally changed deposition of nitrogen", *Forest Eco. and Mgmt.*, **71**, 163-169

Wunderlin, D.A., Diaz, M.D.P., Ame, M.V., Pesce, S.F., Hued, A.C. and Bistoni, M.D.L.A. (2001). "Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality: A case study: Suquia river basin (Cordoba-Argentina)", *Water Research*, **35**(12), 2881-2894

Xu, K., Chen S., Zhao Y., Wang Z., Zhang J., Hayashi S., Murakami S., and Watanabe M. (2005). "Simulated sediment flux during 1998 big-flood of the Yangtze (Changjiang) River, China", *J. of Hydro*., **313**(3 - 4),221-233

Xu, Z. X., and Li, J. Y. (2003). "Estimating basin evapotranspiration using distributed hydrologic model", *J. Hydro. Engineering*, **8** (2), 74-80

Yoshimura, C., Omura, T., Furumai, H. and Tockner, K. (2005). "Present state of rivers and streams in Japan", *River Res. and Applic.*, **21**, 93-112

Yoshino, F., Yoshitani, J. and Horicuhi, T. (1991). "Development of the distributed models for river basins", *Civil Engineering Technical Rep*., Tsukuba, Japan, **32** (10), 54-59 (in Japanese)

Young, R.A., Onstad, C.A., Bosch, D.D., Anderson, W.P. (1989). "Agricultural non-point source pollution model for evaluating agricultural watersheds", *Journal of Soil and Water Conservation,* **44** (2), 168-173.

Zhang, H., and Yamada., K. (1996). "Estimation for urban runoff quality modeling", *Water Sci. Technol.*, **34**(3-4), 49-54.

Zhou, M. C., Ishidaira, H., Hapuarachchi H. P., Magome, J. Kiem, A. S., and Takeuchi, K. (2006). "Estimating potential evapotranspiration using Shuttleworth-Wallace model and NOAA-AVHRR NDVI data to feed a distributed hydrological model over the Mekong River basin", *J. Hydro.,***327**,151-173

APPENDIX A

Appendix A-(1): Description of a Landsat ETM+ image of 2000-05-05 used in this study

Scene ID: LE7110036000012650
WRS Path-Row: 110-36
Acquisition Date: 2000/05/05
NW Corner: 35'34'08'N, 134'12'54'E
NE Corner: 35'16'03'N, 136'17'14'E
Center: 34'36'43'N, 135'01'24'E

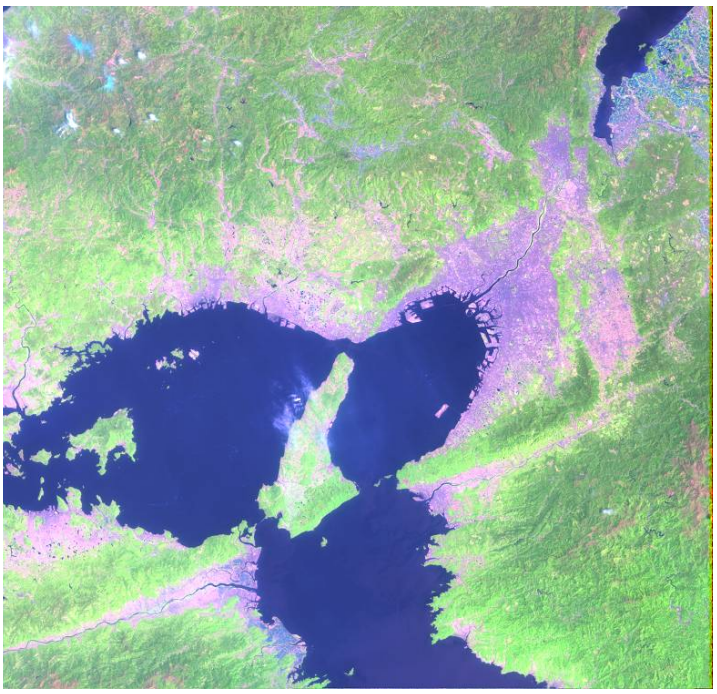
Spacecraft ID: Landsat 7
Sensor ID: ETM+
Receiving Station: SGS
SW Corner: 33'56'39'N, 133'46'43'E
SE Corner: 33'38'55'N, 135'48'40'E

Product Type: Level 1G
Reference Ellipsoid: WGS84
Resampling: NN
Image Quality 1: 9
Flight Path: Descending
Sun Elevation: 62.8316231
Day or Night: Day

Map Projection: UTM
Zone: +53
Cloud Cover: 23
Image Quality 2: 9

Sun Azimuth: 126.7045517

Data Provider: USGS EROS data center



Appendix A-(2): Description of a Landsat ETM+ image of 2000-12-08 used in this study

Scene ID: LE7109036000034350
WRS Path-Row: 109-36
Acquisition Date: 2000/12/08
NW Corner: 35'33'57'N, 135'48'44'E
NE Corner: 35'15'53'N, 137'53'06'E
Center : 34'36'43'N, 136'37'14'E

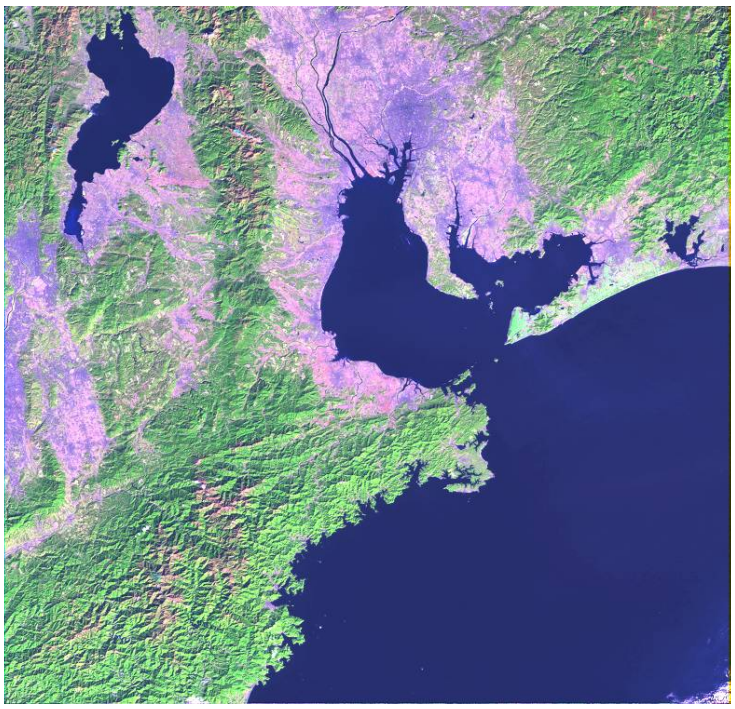
Spacecraft ID: Landsat 7
Sensor ID: ETM+
Receiving Station: SGS
SW Corner: 33'56'20'N, 135'22'24'E
SE Corner: 33'38'37'N, 137'24'23'E

Product Type: Level 1G
Reference Ellipsoid: WGS84
Resampling: NN
Image Quality 1: 9
Flight Path: Descending
Sun Elevation: 29.0830059
Day or Night: Day

Map Projection: UTM
Zone: +53
Cloud Cover: 0
Image Quality 2: 9

Sun Azimuth: 157.093689

Data Provider: GLCF(http://www.landcover.org)



Appendix A-(3): Description of a Landsat ETM+ image of 2000-12-15 used in this study

Scene ID: LE7110036000035050
WRS Path-Row: 110-36
Acquisition Date: 2000/12/15
NW Corner: 35'34'12'N, 134'15'01'E
NE Corner: 35'16'07'N, 136'19'23'E
Center: 34'36'43'N, 135'03'30'E

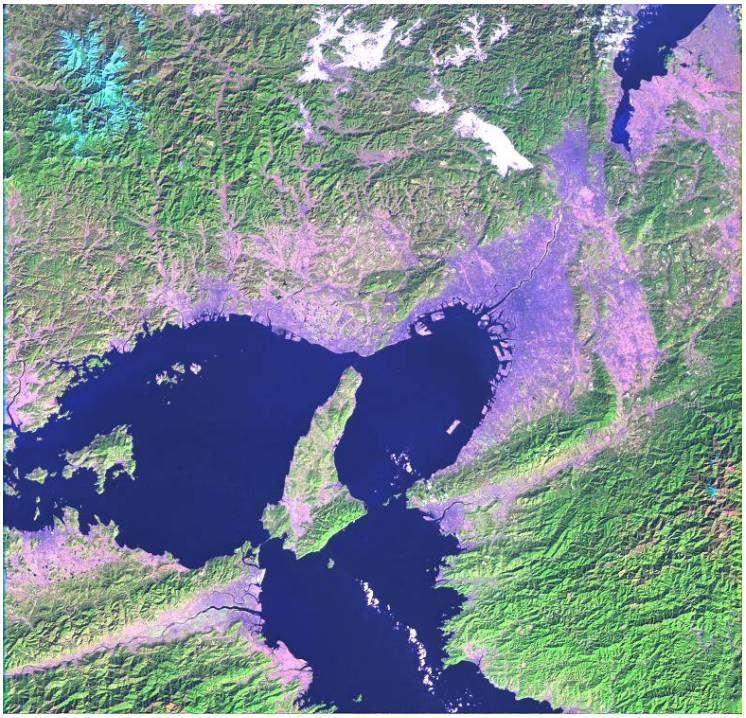
Spacecraft ID: Landsat 7
Sensor ID: ETM+
Receiving Station: SGS
SW Corner: 33'56'37'N, 133'48'44'E
SE Corner: 33'38'53'N, 135'50'43'E

Product Type: Level 1G
Reference Ellipsoid: WGS84
Resampling: NN
Image Quality 1: 9
Flight Path: Descending
Sun Elevation: 28.3146591
Day or Night: Day

Map Projection: UTM
Zone: +53
Cloud Cover: 23
Image Quality 2: 9

Sun Azimuth: 156.5203857

Data Provider: USGS EROS data center



Appendix A-(4): Description of a Landsat ETM+ image of 2001-10-15 used in this study

Scene ID: LE7110036000128850
WRS Path-Row: 110-36
Acquisition Date: 2001/10/15
NW Corner: 35'34'12'N, 134'13'56'E
NE Corner: 35'16'05'N, 136'18'20'E
Center : 34'36'43'N, 135'02'23'E

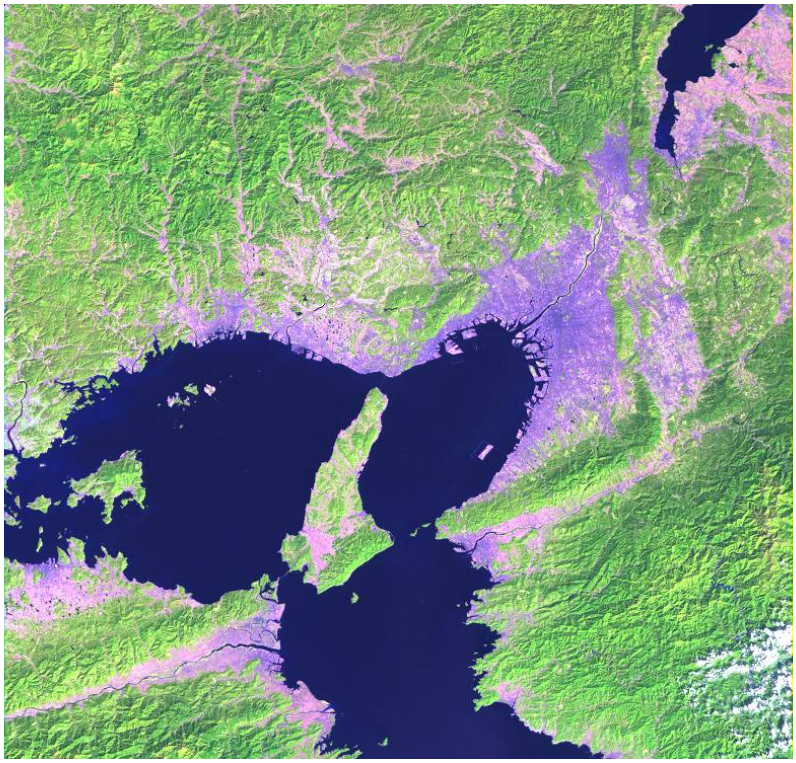
Spacecraft ID: Landsat 7
Sensor ID: ETM+
Receiving Station: SGS
SW Corner: 33'56'29'N, 133'47'33'E
SE Corner: 33'38'42'N, 135'49'35'E

Product Type: Level 1G
Reference Ellipsoid: WGS84
Resampling: NN
Image Quality 1: 9
Flight Path: Descending
Sun Elevation: 42.6616096
Day or Night: Day

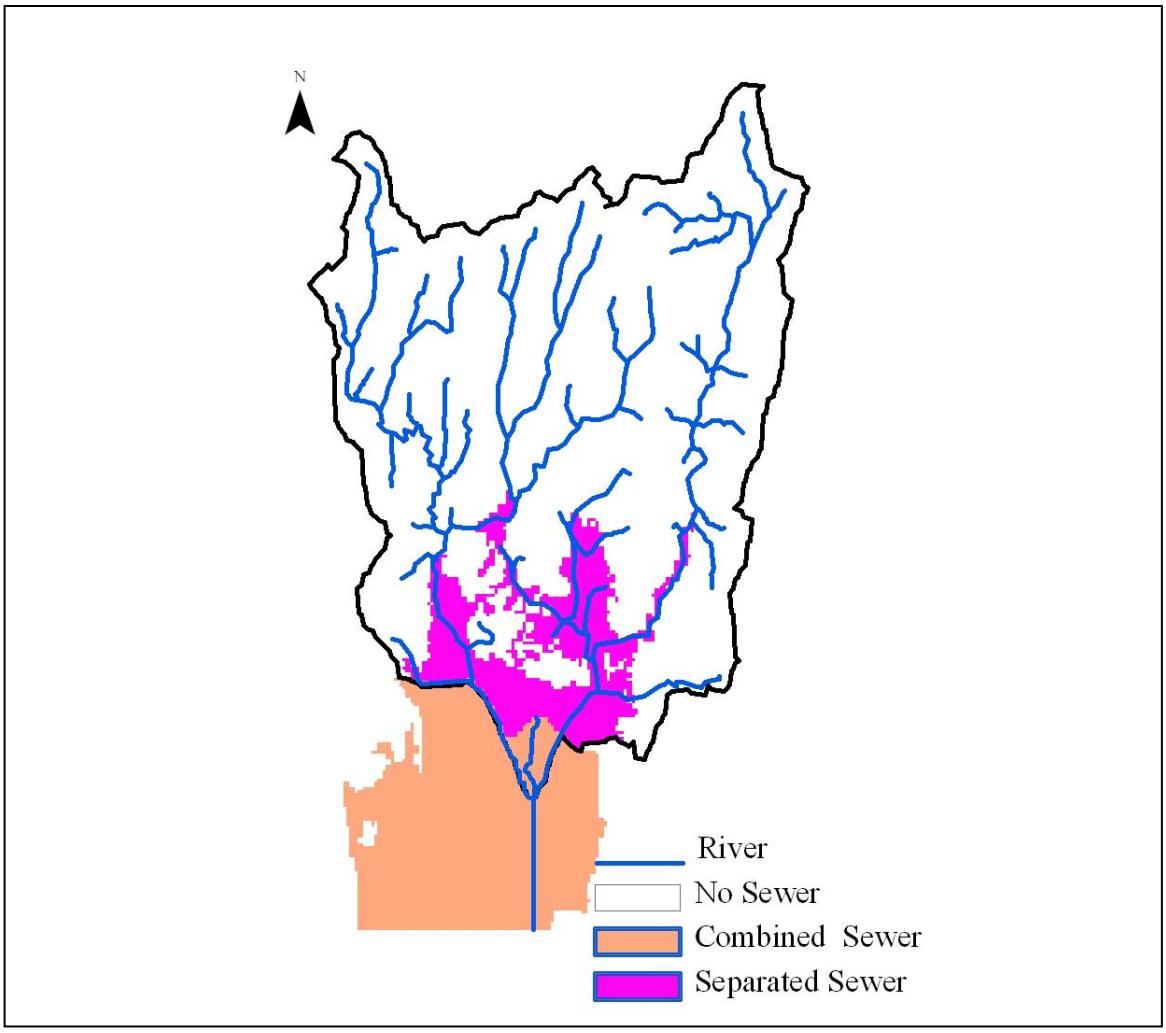
Map Projection: UTM
Zone: +53
Cloud Cover: 0
Image Quality 2: 9

Sun Azimuth: 151.6215515

Data Provider: GLCF(http://www.landcover.org)

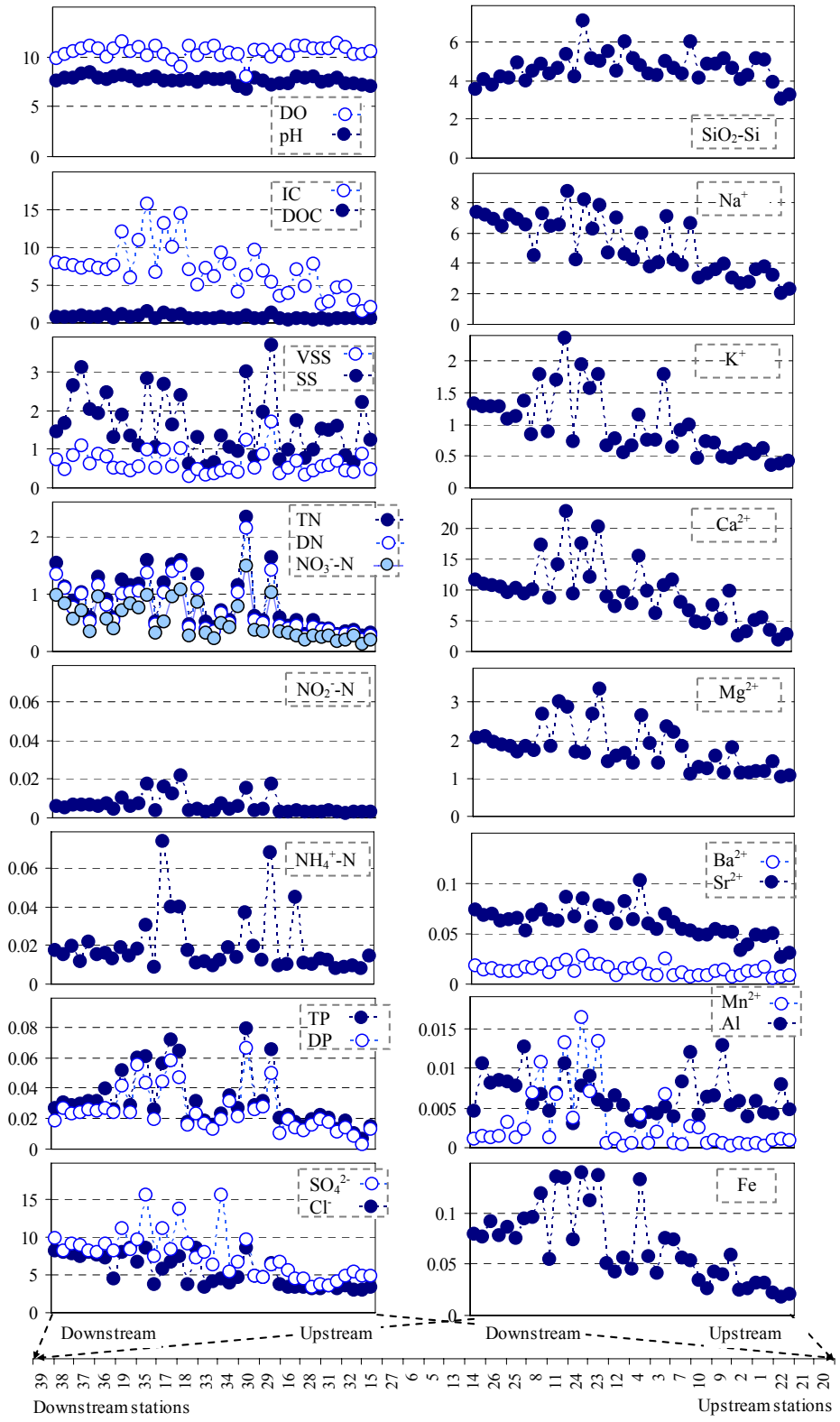


Appendix A-(5): Map of the study area showing the separated and combined sewer areas

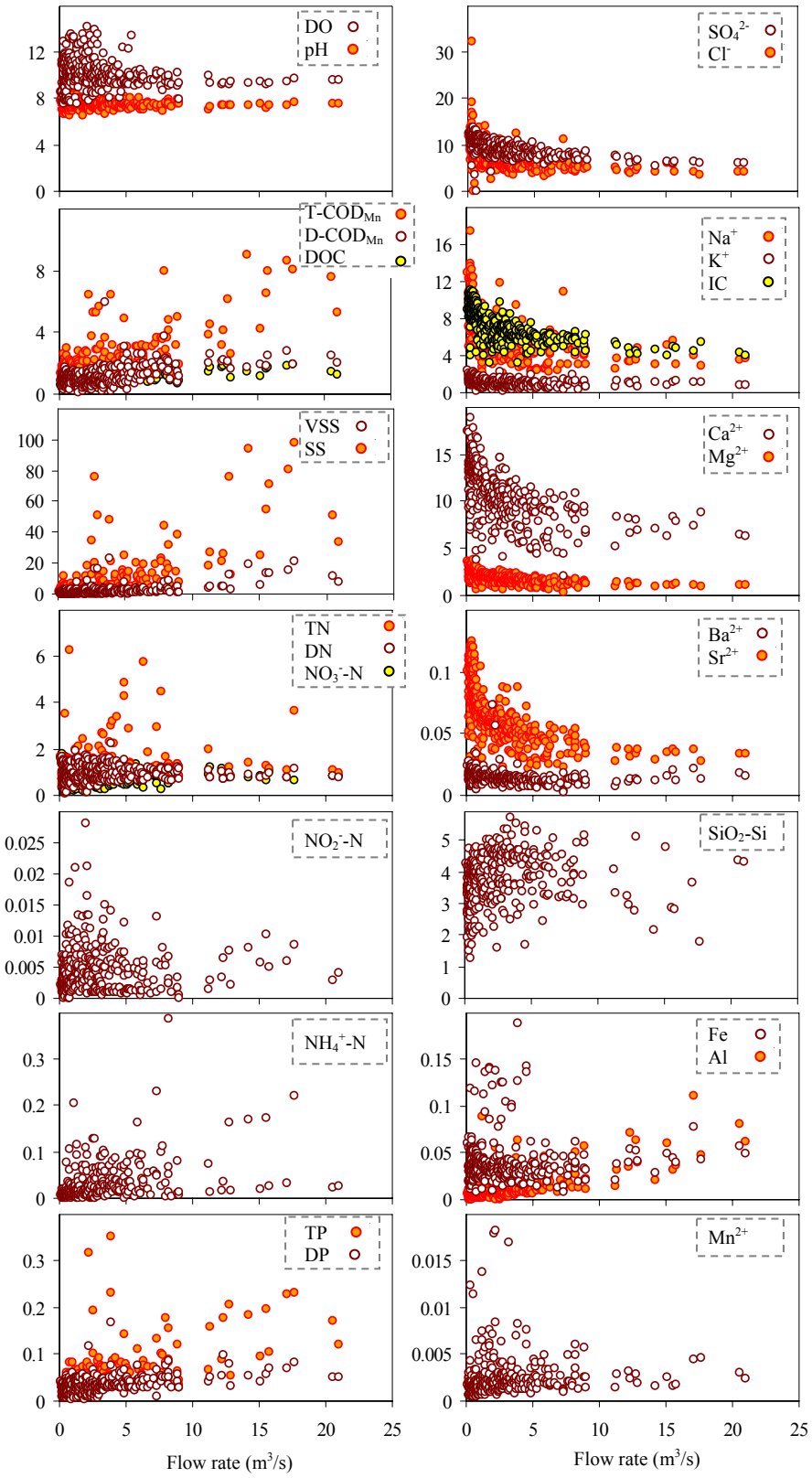


APPENDIX B

Appendix B-(1): Concentration differences of water quality indices (WQIs) at different stations (except pH, unit of all other WQIs was mg/L)



Appendix B-(2): Concentration differences of water quality indices (WQIs) compared with flow rates (m^3/s) observed at different temporal conditions (except pH, unit of all other WQIs was mg/L)



Appendix B-(3): Correlation among WQIs for spatial dataset

	$T\text{-}COD_{Mn}$	$D\text{-}COD_{Mn}$	DOC	IC	SS	VSS	TN	DN	NO_2^--N	NO_3^--N	NH_4^+-N	TP	DP	Cl^-	SO_4^{2-}	Na^+	K^+	Mg^{2+}	Ca^{2+}	Sr^{2+}	Ba^{2+}	SiO_2	Fe	Al	Mn^{2+}	SFR
$T\text{-}COD_{Mn}$	**1.00**																									
$D\text{-}COD_{Mn}$	**0.84**	**1.00**																								
DOC	**0.68**	**0.74**	**1.00**																							
IC	0.48	**0.59**	0.46	**1.00**																						
SS	**0.61**	0.50	0.47	0.12	**1.00**																					
VSS	**0.58**	0.44	0.42	0.07	**0.82**	**1.00**																				
TN	**0.54**	**0.58**	0.49	0.46	0.35	0.35	**1.00**																			
DN	**0.54**	**0.58**	0.48	0.45	0.37	0.39	**0.98**	**1.00**																		
NO_2^--N	**0.54**	**0.54**	0.43	0.49	0.18	0.24	0.46	0.45	**1.00**																	
NO_3^--N	0.46	0.50	0.38	0.45	0.22	0.29	**0.87**	**0.89**	**0.51**	**1.00**																
NH_4^+-N	0.50	0.47	0.47	0.27	0.29	0.36	0.45	0.43	0.47	0.27	**1.00**															
TP	**0.59**	**0.59**	0.43	**0.54**	0.32	0.32	**0.74**	**0.73**	**0.58**	**0.69**	0.41	**1.00**														
DP	**0.52**	**0.53**	0.37	**0.55**	0.20	0.25	**0.65**	**0.67**	**0.57**	**0.70**	0.33	**0.91**	**1.00**													
Cl^-	**0.50**	**0.57**	**0.52**	0.41	0.32	0.25	**0.60**	**0.58**	0.32	**0.60**	0.18	0.45	0.44	**1.00**												
SO_4^{2-}	0.43	**0.58**	0.47	**0.73**	0.21	0.09	**0.52**	**0.53**	0.38	**0.52**	0.13	0.45	0.41	**0.52**	**1.00**											
Na^+	**0.54**	**0.65**	**0.54**	**0.67**	0.22	0.16	**0.59**	**0.57**	0.49	**0.62**	0.20	**0.52**	**0.53**	**0.87**	**0.70**	**1.00**										
K^+	**0.62**	**0.67**	**0.61**	**0.71**	0.28	0.29	**0.58**	**0.59**	**0.60**	**0.63**	0.27	**0.62**	**0.67**	**0.63**	**0.66**	**0.75**	**1.00**									
Mg^{2+}	0.35	0.45	0.38	**0.82**	0.07	0.01	**0.55**	**0.55**	0.39	**0.55**	0.14	**0.56**	**0.56**	0.49	**0.71**	**0.62**	**0.62**	**1.00**								
Ca^{2+}	0.49	**0.60**	**0.56**	**0.95**	0.17	0.11	**0.54**	**0.53**	0.46	**0.51**	0.27	**0.51**	**0.50**	0.49	**0.81**	**0.70**	**0.73**	**0.85**	**1.00**							
Sr^{2+}	0.19	0.28	0.22	**0.63**	-0.08	-0.15	0.29	0.26	0.17	0.35	0.00	0.22	0.29	0.44	**0.69**	**0.67**	0.49	**0.59**	**0.68**	**1.00**						
Ba^{2+}	0.37	0.37	0.38	**0.54**	0.09	0.06	0.42	0.41	0.38	0.39	0.28	0.38	0.39	0.24	0.49	0.44	**0.52**	**0.51**	**0.61**	0.42	**1.00**					
SiO_2	0.18	0.19	0.12	0.32	0.01	0.02	0.20	0.19	0.26	0.13	0.38	0.29	0.25	-0.16	0.15	0.10	-0.01	0.10	0.27	0.12	0.39	**1.00**				
Fe	0.23	0.28	-0.10	0.46	-0.01	0.00	0.15	0.17	0.39	0.28	-0.04	0.36	0.41	0.04	0.40	0.24	0.32	0.33	0.33	0.15	0.15	0.26	**1.00**			
Al	0.34	0.33	0.25	0.06	0.36	0.27	0.20	0.18	-0.12	0.00	0.19	0.12	0.01	0.15	0.11	0.10	0.03	0.05	0.09	0.01	0.07	0.11	-0.03	**1.00**		
Mn^{2+}	**0.53**	**0.56**	**0.51**	**0.64**	0.20	0.16	0.42	0.42	**0.50**	0.40	0.37	0.48	0.48	0.30	**0.57**	0.48	**0.58**	**0.53**	**0.66**	0.37	**0.55**	0.32	0.28	0.05	**1.00**	
SFR#	-0.03	-0.12	-0.23	-0.38	0.12	0.21	0.02	0.01	0.01	-0.03	0.13	-0.07	-0.11	-0.23	-0.44	-0.37	-0.34	-0.37	-0.39	**-0.55**	-0.21	0.14	-0.03	0.00	-0.22	**1.00**

#Specific flow rate

Appendix B-(4): Correlation among WQIs for temporal dataset

	T-COD_{Mn}	D-COD_{Mn}	DOC	IC	SS	VSS	TN	DN	NO_2^--N	NO_3^--N	NH_4^+-N	TP	DP	Cl^-	SO_4^{2-}	Na^+	K^+	Mg^{2+}	Ca^{2+}	Sr^{2+}	Ba^{2+}	SiO_2	Fe	Al	Mn^{2+}	SFR
T-COD_{Mn}	**1.00**																									
D-COD_{Mn}	**0.63**	**1.00**																								
DOC	**0.66**	**0.59**	**1.00**																							
IC	-0.28	-0.16	-0.30	**1.00**																						
SS	**0.90**	0.49	**0.58**	-0.29	**1.00**																					
VSS	**0.86**	0.42	**0.56**	-0.27	**0.93**	**1.00**																				
TN	0.01	-0.01	0.01	-0.01	0.02	0.02	**1.00**																			
DN	0.02	-0.10	0.04	-0.01	0.00	0.07	0.25	**1.00**																		
NO_2^--N	0.17	0.05	0.32	-0.20	0.16	0.26	0.02	0.09	**1.00**																	
NO_3^--N	-0.05	-0.10	-0.12	0.19	-0.09	-0.06	-0.01	**0.60**	-0.11	**1.00**																
NH_4^+-N	0.43	0.20	0.28	-0.21	0.35	**0.52**	0.07	0.18	0.27	-0.11	**1.00**															
TP	**0.82**	0.49	**0.60**	-0.33	**0.74**	**0.77**	0.08	0.19	0.28	-0.01	**0.66**	**1.00**														
DP	**0.61**	0.42	**0.54**	-0.30	0.49	**0.56**	0.09	0.36	0.24	0.12	**0.59**	**0.85**	**1.00**													
Cl^-	-0.29	-0.24	-0.28	0.44	-0.28	-0.25	-0.02	0.21	0.02	0.44	-0.18	-0.29	-0.24	**1.00**												
SO_4^{2-}	-0.38	-0.30	-0.34	**0.62**	-0.38	-0.35	-0.01	0.31	-0.09	**0.58**	-0.26	-0.36	-0.30	**0.66**	**1.00**											
Na^+	-0.24	-0.20	-0.34	**0.61**	-0.25	-0.21	-0.13	0.11	-0.05	0.44	-0.19	-0.25	-0.22	**0.76**	**0.63**	**1.00**										
K^+	0.05	0.02	-0.03	**0.55**	0.01	0.01	-0.09	0.24	0.04	**0.56**	-0.13	0.02	0.05	**0.54**	**0.56**	**0.81**	**1.00**									
Mg^{2+}	-0.34	-0.29	-0.36	**0.63**	-0.30	-0.26	-0.12	0.12	-0.05	0.43	-0.28	-0.35	-0.33	**0.53**	**0.67**	**0.85**	**0.79**	**1.00**								
Ca^{2+}	-0.25	-0.22	-0.29	**0.66**	-0.22	-0.19	-0.16	0.03	-0.02	0.34	-0.23	-0.27	-0.26	**0.51**	**0.63**	**0.85**	**0.80**	**0.93**	**1.00**							
Sr^{2+}	-0.27	-0.27	-0.23	0.43	-0.24	-0.17	-0.10	-0.08	0.23	0.13	-0.17	-0.28	-0.32	0.40	0.46	**0.58**	**0.51**	**0.67**	**0.65**	**1.00**						
Ba^{2+}	0.10	-0.03	-0.02	0.35	0.03	0.12	-0.08	0.08	0.21	0.09	0.25	0.19	0.09	0.20	0.30	0.41	0.42	0.44	**0.52**	0.38	**1.00**					
SiO_2	-0.17	-0.07	-0.09	-0.27	-0.20	-0.20	0.09	0.23	-0.17	0.13	-0.11	-0.04	0.12	-0.23	-0.12	-0.26	-0.27	-0.23	-0.30	-0.31	-0.23	**1.00**				
Fe	0.33	0.02	0.25	-0.07	0.22	0.36	0.01	0.08	0.19	-0.01	0.33	0.40	0.38	-0.13	-0.09	0.07	0.16	0.05	0.15	0.10	0.33	-0.04	**1.00**			
Al	**0.57**	0.43	**0.52**	-0.37	**0.55**	**0.51**	0.25	-0.01	0.17	-0.20	0.28	**0.62**	**0.57**	-0.36	-0.43	-0.36	-0.16	-0.38	-0.31	-0.27	0.03	0.03	0.29	**1.00**		
Mn^{2+}	0.17	0.10	0.27	-0.21	0.14	0.21	0.25	0.11	0.23	-0.15	0.28	0.29	0.31	-0.24	-0.22	-0.23	-0.09	-0.17	-0.15	-0.11	0.15	0.12	0.43	0.45	**1.00**	
SFR[#]	**0.72**	**0.51**	**0.55**	**-0.63**	**0.73**	**0.65**	0.01	0.01	0.07	-0.12	0.27	**0.63**	0.49	-0.45	**-0.60**	**-0.52**	-0.32	**-0.58**	**-0.57**	-0.47	-0.23	0.12	0.06	**0.66**	0.14	**1.00**

#Specific flow rate

APPENDIX C

Appendix C-(1): Pictures of forest areas showing three main vegetation types (Deciduous, Evergreen, and Mixed) taken inside the study area (Date: 2006/12/16; Place: Near Kurama mountain area; Approximate position: $35^0$06'51"N $135^0$46'21"E)